

Primjena sustava za računalno potpomognuto prevođenje u domeni obrazovanja

Matijašec, Ksenija

Master's thesis / Diplomski rad

2023

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

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:131:307713>

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

Download date / Datum preuzimanja: **2024-07-28**



Sveučilište u Zagrebu
Filozofski fakultet
University of Zagreb
Faculty of Humanities
and Social Sciences

Repository / Repozitorij:

[ODRAZ - open repository of the University of Zagreb
Faculty of Humanities and Social Sciences](#)



SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI
SMJER INFORMATIKA NASTAVNIČKI
Ak. god. 2022./2023.

KSENIJA MATIJAŠEC

**Primjena sustava za računalno potpomognuto prevođenje
u domeni obrazovanja**

Diplomski rad

Mentor: prof. dr. sc. Sanja Seljan

Zagreb, rujan 2023.

Izjava o akademskoj čestitosti

Izjavljujem da je ovaj rad rezultat mog vlastitog rada koji se temelji na istraživanjima te objavljenoj i citiranoj literaturi. Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava. Također izjavljujem da nijedan dio rada nije korišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

Ksenija Matijašec (potpis)

Sadržaj

Sadržaj.....	i
1. Uvod.....	1
2. Razvoj sustava za prevođenje.....	3
3. Računalno potpomognuto prevođenje.....	7
3.1. Prijevodna memorija.....	9
3.1.1. Segmentacija.....	10
3.1.2. Sravnjivanje.....	12
3.1.3. Podudaranje.....	15
3.2. Terminološke baze.....	17
3.2.1. Izrada terminoloških baza.....	17
3.2.2. Integracija terminoloških baza u CAT sustave.....	19
3.2.3. Metapodaci u terminološkim bazama.....	20
3.3. Procjena kvalitete prijevoda.....	21
3.3.1. Osiguravanje kvalitete.....	22
3.3.2. Sustavi za upravljanje prevođenjem.....	27
3.4. Integracija strojnog prevođenja u CAT sustave.....	28
4. Istraživanje.....	31
4.1. Resursi i alati.....	32
4.1.1. Stvaranje prijevodne memorije.....	32
4.1.2. Stvaranje terminološke baze.....	35
4.1.3. Integracija strojnog prevođenja.....	35
4.2. Testiranje sustava.....	36
4.2.1. Analiza izvješća alata Trados Studio.....	36
4.2.2. Proces prevođenja.....	37
4.2.3. Osiguravanje kvalitete prijevoda.....	39

4.3. Usporedna analiza prijevoda	42
4.3.1. Kriteriji	44
4.3.2. Analiza rezultata	45
5. Zaključak	57
6. Literatura:.....	59
Popis slika.....	64
Popis tablica.....	65
Popis grafova	66
Sažetak.....	1
Summary.....	2

1. Uvod

Potreba za primjenom tehnologije u komunikaciji danas je sve izraženija. To je posebno izraženo zbog rastuće višejezičnosti, korištenja pametnih uređaja i potrebe za brzim pristupom informacijama na različitim jezicima. U tom kontekstu, prevođenje pomoću tehnologije postaje ključno za poboljšanje produktivnosti i opće kvalitete prijevoda (Seljan i sur., 2007).¹ Nadalje, industrija prevođenja se suočava s izazovima poput manjka prevoditelja i pritiska vezanih uz vremenske rokove. Uz to, promjene su potaknute čimbenicima kao što su višejezičnost, jezična tehnologija i razvoj interneta, što dovodi do postavljanja novih standarda za kvalitetu, učinkovitost i dosljednost u prevođenju (Kučiš i sur., 2009).² Razlog prevođenja općenito je prijenos informacija i znanja s jednog jezika na drugi stoga se razvijaju različiti sustavi za prevođenje. Sustavi za računalno potpomognuto prevođenje (engl. *Computer Assisted Translation, Computer Aided Translation, CAT*) su ključni u olakšavanju međunarodne komunikacije i razmjene informacija te je njihov cilj pružiti podršku korisnicima tijekom cijelog procesa prevođenja, olakšavajući im rad i rezultirajući kvalitetnijim prijevodima (Folaron, 2010)³ u institucijama, kao što je Europska komisija, ali i u privatnim tvrtkama (Seljan, 2018).⁴ U novije vrijeme, posebnu pažnju su izazvali alati umjetne inteligencije, kao što je ChatGPT koji, osim ostalih funkcionalnosti nudi i funkcionalnost automatskog strojnog prevođenja, ali koji zahtijeva naknadno uređivanje i provjeru kvalitete (Dunđer i sur., 2021⁵; Seljan i sur., 2020⁶, Dunđer i sur. 2020⁷). Osim toga, dostupni su drugi brojni online alati za prevođenje koji se mogu koristiti samostalno, s CAT alatima ili npr. alatima za konkordancije (Jaworski i sur., 2021)⁸, u različitim domenama.

¹ Seljan, S., Gašpar, A. & Pavuna, D. (2007). Sentence Alignment as the Basis for Translation Memory Database. INFUTURE2007: "Digital Information and Heritage". 299-311. str. 301

² Kučiš, V., Seljan, S., Klasnić, K. (2009). Evaluation of electronic translation tools through quality parameters. INFUTURE 2009: Digital Resources and Knowledge Sharing. 341-351. str. 342

³ Folaron, D. A. (2010). Translation tools. Y. Gambier, & L. van Doorslaer (Ur.), Handbook of Translation Studies. 1. 429-436. str. 429

⁴ Seljan, S. (2018). Total Quality Management Practice in Croatian Language Service Provider Companies. ENTRENOVA '18 - ENTERPRISE RESEARCH INNOVATION CONFERENCE. 4 (1). 461-469. str. 431

⁵ Dunđer, I., Seljan, S., Pavlovski, M. (2021). What Makes Machine-Translated Poetry Look Bad? A Human Error Classification Analysis. Central European Conference on Information and Intelligent Systems. 183-191.

⁶ Seljan, S., Dunđer, I., Pavlovski, M. (2020). Human Quality Evaluation of Machine-Translated Poetry. MIPRO, 1040-1045.

⁷ Dunđer, I., Seljan, S., Pavlovski, M. (2020). Automatic Machine Translation of Poetry and a Low-Resource Language Pair. MIPRO, 1034-1039.

⁸ Jaworski, R., Duđer, I., Seljan, S. (2021). Usability Analysis of the Concordia Tool Applying Novel Concordance Searching. Information Technology and Systems. 1, 128-138. str. 129

Ovaj rad istražuje funkcionalnosti CAT sustava te njegovu primjenu. Cilj istraživanja je ispitati učinkovitost primjene CAT alata radi utvrđivanja raznolikosti stručne terminologije, vrsta i kategorija pogrešaka i stvaranja prijevoda sukladnog referentnom prijevodu.

U teorijskom dijelu rada prikazuje se razvoj sustava za prevođenje s fokusom na računalno potpomognuto prevođenje. Nadalje, detaljno se opisuju ključne komponente računalno potpomognutog prevođenja u koje spadaju prijevodne memorije, terminološke baze, osiguravanje kvalitete i integracija automatskog strojnog prevođenja. U sklopu ovog dijela prikazuju se prednosti i izazovi koji proizlaze iz upotrebe navedenih komponenti te se prikazuju načini na koje ta tehnologija može poboljšati produktivnost i preciznost prevođenja.

U praktičnom dijelu rada provedeno je pilot istraživanje temeljeno na usporedbi tradicionalnog prevođenja i prevođenja primjenom CAT alata. Za potrebe pilot istraživanja, kreirani su resursi koji su integrirani u CAT alat, a to su prijevodna memorija, terminološka baza i integrirano je automatsko strojno prevođenje. Ovo istraživanje se fokusiralo na to kako pripremljeni resursi utječu na proces prevođenja pomoću CAT alata. Na kraju su prikazani rezultati istraživanja temeljeni na potrebnom vremenu za prevođenje, vrsti i kategoriji pogrešaka, stvaranju prijevoda sukladnog referentnom prijevodu i korištenju stručne terminologije. Na rezultate pilot istraživanja utjecala je ograničena količina pripremljenih resursa i mali broj ispitanika te sami alati korišteni tijekom provedbe istraživanja. Zbog toga, ovo pilot istraživanje se može smatrati samo početnim korakom u razumijevanju potencijalnih prednosti i nedostataka u korištenju CAT alata.

2. Razvoj sustava za prevođenje

Prema Hutchinsu (1995)⁹ koncept prevođenja korištenjem stroja prvi put se pojavljuje u 17. stoljeću pojavom ideje univerzalnog jezika. Ovaj koncept je uključivao primjenu mehaničkih rječnika kako bi se prevladale jezične prepreke. Prvi konkretni prijedlozi i eksperimenti pojavljuju se tek u 20. stoljeću te predstavljaju početke strojnog prevođenja. Strojno prevođenje (engl. *Machine Translation*, MT) predstavlja računalni softver specifično stvoren za prevođenje teksta s jednog jezika na drugi, bez ikakvog sudjelovanja ljudi u procesu prevođenja. Glavni cilj strojnog prevođenja jest razviti sustav koji može efikasno prevesti tekst s izvornog jezika na ciljni jezik, osiguravajući da prevedeni tekst zadržava isto značenje kao i tekst koji se prevodi (Irfanu, 2017).¹⁰ Folaron (2010)¹¹ sistematizira i objašnjava tehnologije prevođenja, tako predstavljajući strojno prevođenje, ljudski potpomognuto strojno prevođenje (engl. *Human Aided Machine Translation*, HAMT), strojno potpomognuto ljudsko prevođenje (engl. *Machine Aided Human Translation*, MAHT) i računalno potpomognuto prevođenje (engl. *Computer Assisted Translation*, *Computer Aided Translation*, CAT). Strojno prevođenje predstavljaju automatski prijevodi teksta bez ikakve ljudske intervencije. Ljudski potpomognuto strojno prevođenje je automatsko prevođenje koje podržava čovjeka, a strojno potpomognuto ljudsko prevođenje obuhvaća programe koji imaju za cilj podržati proces prevođenja korištenjem različitih potrebnih sustava, ali koji se i dalje uglavnom oslanjaju na ljudski rad. Sustavi računalno potpomognutog prevođenja spadaju u kategoriju strojno potpomognutog ljudskog prevođenja i smatraju se najrazvijenijim i najčešćim sustavima na tržištu prevođenja.

Hutchins (1986)¹² opisuje rane primjere sustava koji se temelje na strojevima za prijevod te se oni oslanjaju na jezična pravila i rječnike za prevođenje teksta. Ti rani sustavi uključivali su ručno kodiranje gramatike i jezičnih pravila, koja su često bila složena i ograničena na određene jezične parove. Nadalje, Dovedan i suradnici (2002)¹³ pružaju prve primjere strojnog prevođenja. Ti primjeri su opisani u patentima koji su neovisno izdani 1933. godine, od strane

⁹ Hutchins, J. (1995). *Machine Translator: A Brief History*. E. F. K. Koerner & R. E. Asher (Ed.). *Concise History of the Language Sciences*. 431-445. str. 433

¹⁰ Irfan, M. (2017). *Machine Translation*. Department of Computer Science. Bahria University Islamabad. str. 1

¹¹ Folaron, D. A. (2010). *Translation tools*. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 429-436. str. 433

¹² Hutchins, J. (1986). *Machine Translation: Past, Present, Future*. Chichester: Ellis Horwood. str. 24-28

¹³ Dovedan, Z., Seljan, S. & Vučković K. (2002). *Strojno prevođenje kao pomoć u procesu komunikacije*. *Informatologia*. 35 (4). 283-291. str. 284

Georgea Artsrounija i Petera Smirnova Troyanskog. Uređaj za pohranu na papirnatu traku koji je osmislio Artsrouni omogućavao je pronalaženje ekvivalenta bilo koje riječi na drugom jeziku. Troyanskijev patent smatra se značajnim i sastoji se od tri faze mehaničkog prevođenja. Prva faza uključuje osobu koja poznaje samo izvorni jezik i pretvara riječi u njihove osnovne oblike. Drugu fazu obavlja stroj koji prevodi osnovne oblike iz prvog jezika u osnovne oblike drugog jezika. Treću fazu izvodi osoba koja poznaje samo ciljni jezik i prevodi osnovne oblike koje je stroj proizveo u vlastiti jezik. Hutchins (1995)¹⁴, osim Artsrounija i Troyanskog, navodi kako je jedan od važnih začetnika računalnog prevođenja bio Warren Weaver. On je najpoznatiji po svom utjecajnom memorandumu pod nazivom "Prijevod", objavljenom 1949. godine. U tom memorandumu, Weaver je predložio ideju korištenja računala za automatsko prevođenje ljudskih jezika te daje konkretne prijedloge rješavanja problema dvosmislenosti. Naglašavao je izazove i mogućnosti strojnog prevođenja te isticao važnost razlaganja jezika na osnovne elemente, analizu njihove strukture i rekonstrukciju na ciljnom jeziku.

Dovedan i suradnici (2002)¹⁵ ističu da su nakon prvih sustava za prevođenje, tijekom dugog niza godina, sustavi prevođenja uglavnom temeljeni na izravnim prijevodima pomoću dvojezičnih rječnika. Ti sustavi su se oslanjali na relativno ograničenu analizu sintaktičkih struktura. Unutar tog razdoblja prevladavala je ideja stvaranja savršenog automatskog strojnog prevođenja. Međutim, tijekom 1980-ih napredak u računalnoj lingvistici omogućio je mnogo sofisticiranije pristupe i razvoj sustava temeljenih na pravilima (eng. *Rule-based Machine Translation systems*, RBMT) (Hutchins, 2003).¹⁶ Kod sustava temeljenim na pravilima razvijaju se gramatička pravila i leksikoni za analizu i prijevod rečenica (Irfan, 2017).¹⁷ Sustav koristi navedene elemente kako bi analizirao rečenice na izvornom jeziku i preveo ih. Sustavi temeljeni na pravilima zahtijevaju velik napor u pripremi pravila i jezičnih resursa, ali omogućuju kontrolu nad prijevodima (Sreelekha, 2017).¹⁸ U tim naprednijim sustavima, tekstovi na izvornom jeziku analiziraju se u apstraktne prikaze "značenja". Ova analiza uključuje programe za identifikaciju strukture riječi (morfologija) i strukture rečenica (sintaksa), kao i probleme dvosmislenosti (semantika). Cilj je postići nedvosmislene apstraktne

¹⁴ Hutchins, J. (1995). Machine Translator: A Brief History. E. F. K. Koerner & R. E. Asher (Ed.). Concise History of the Language Sciences. 431-445. str. 433

¹⁵ Dovedan, Z., Seljan, S. & Vučković K. (2002). Strojno prevođenje kao pomoć u procesu komunikacije. Informatologia. 35 (4). 283-291. str 285

¹⁶ Hutchins, J. (2003). The development and Use of Machine Translation Systems and Computerbased Translation Tools. International Journal of Translation. 15(1). 6-26. str. 7

¹⁷ Irfan, M. (2017). Machine Translation. Department of Computer Science. Bahria University Islamabad. str. 1

¹⁸ Sreelekha, S. (2017). Statistical Vs Rule Based Machine Translation: A Case Study on Indian Language Perspective. Springer Journal of Advances in Intelligence and Soft Computing. 663-675. str. 664, 665

prikaze koji će služiti kao temelj za generiranje teksta na jednom ili više ciljanih jezika (Irfan, 2017).¹⁹ Ovakav neizravan pristup omogućuje sustavima da dublje razumiju i tumače značenje izvornog teksta, umjesto da se oslanjaju samo na prijevod riječ po riječ. Analizom strukture i značenja, sustavi strojnog prevođenja mogu bolje obuhvatiti kontekstualne informacije izvornog teksta te ih prenijeti na ciljni jezik. Napredak u računalnoj lingvistici i korištenje neizravnih pristupa prevođenju označava važan korak naprijed u razvoju sustava za prevođenje, otvarajući put za sve sofisticiranije i preciznije sustave koji bolje zadovoljavaju zahtjeve prevođenja (Hutchins, 2003).²⁰

Osim razvoja sustava temeljenih na pravilima, 1980-ih se pojavljuju početne ideje o integraciji različitih računalnih sadržaja na jednom mjestu (Dovedan i sur., 2002).²¹ U tom razdoblju razvijaju se elektroničke terminološke baze, sve je veća potreba za pridržavanjem terminoloških standarda, vrši se mnoštvo prijevoda i pojavio se napredak u korištenju prethodnih primjera prijevoda. Stoga su prednosti osobnih računala i obrade teksta sve naglašenije (Hutchins, 2003).²² Razdoblje 1980-ih pridonijelo je najvećem napretku u području prevođenja pojavom prevoditeljskih radnih stanica. Prema Hutchinsu (2003)²³ prevoditeljske radne stanice temelje se na centralnoj prevoditeljskoj radnoj stanici koja uključuje različite dodatne računalne sadržaje kao što su terminološke baze, mogućnost izrade tekstualnog glosara, pristup prethodnim prijevodima iz strojnog prevođenja, prijevodno pamćenje i postupci osiguravanja kvalitete. Svi ovi resursi i mogućnosti su integrirani u radnu stanicu korisnika, omogućavajući korisniku da odabere alate koje želi koristiti i s kojima obavlja stvarni prijevod. U ovom kontekstu korisnik ima potpunu kontrolu nad procesom prevođenja i nije podređen stroju.

Ovaj koncept prevoditeljskih radnih stanica omogućuje korisnicima da maksimalno iskoriste prednosti računalne tehnologije i alata za podršku u procesu prevođenja. Integracija različitih funkcionalnosti u jednu radnu stanicu pruža korisnicima okruženje u kojem mogu učinkovito upravljati svojim prijevodnim projektima (Hutchins, 1995).²⁴ Osim toga,

¹⁹ Irfan, M. (2017). Machine Translation. Department of Computer Science. Bahria University Islamabad. str. 1

²⁰ Hutchins, J. (2003). The development and Use of Machine Translation Systems and Computerbased Translation Tools. *International Journal of Translation*. 15(1). 6-26. str. 8

²¹ Dovedan, Z., Seljan, S. & Vučković K. (2002). Strojno prevođenje kao pomoć u procesu komunikacije. *Informatologia*. 35 (4). 283-291. str. 288, 289

²² Hutchins, J. (2003). The development and Use of Machine Translation Systems and Computerbased Translation Tools. *International Journal of Translation*. 15(1). 6-26. str. 8

²³ Hutchins, J. (2003). The development and Use of Machine Translation Systems and Computerbased Translation Tools. *International Journal of Translation*. 15(1). 6-26. str. 14

²⁴ Hutchins, J. (1995). Machine Translator: A Brief History. E. F. K. Koerner & R. E. Asher (Ed.). *Concise History of the Language Sciences*. 431-445. str. 441

prevoditeljske radne stanice ili danas programski paketi, omogućuju bolju organizaciju i upravljanje resursima, što rezultira dosljednošću i produktivnošću u procesu prevođenja. Stoga, razvoj ovakvih alata predstavlja važan korak u integraciji računalne tehnologije i procesa prevođenja, pružajući korisnicima alate, resurse i kontrolu nad prijevodnim projektima (Hutchins, 2003).²⁵ Prema Folaronu (2010)²⁶ ovaj razvoj sustava za prevođenje je pridonio razvoju današnjih sustava za računalno potpomognuto prevođenje i današnjih programskih paketa. Navedeni sustavi su osmišljeni kako bi poboljšali produktivnost i preciznost korisnika pružajući im integrirano okruženje i pristup raznim korisnim alatima.

²⁵ Hutchins, J. (2003). The development and Use of Machine Translation Systems and Computerbased Translation Tools. *International Journal of Translation*. 15(1). 6-26. str. 14-15

²⁶ Folaron, D. A. (2010). Translation tools. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 429-436. str. 435

3. Računalno potpomognuto prevođenje

Prema Bowkeru i Fisheru (2010)²⁷ računalno potpomognuto prevođenje (engl. *Computer Assisted Translation*, *Computer Aided Translation*, CAT) predstavlja upotrebu računalnog softvera koji pomaže korisniku u procesu prevođenja gdje korisnik zadržava konačnu odgovornost za prijevod. Razvoj CAT sustava pratio je razvoj računalne i mrežne industrije. U počecima razvoja postojali su kao samostalni alati koji se koriste na pojedinim računalima. Zatim se razvijaju klijentski i poslužiteljski alati koji se koriste unutar mreža tvrtki, što je omogućilo suradnju između timova i centralizirano upravljanje resursima (Hutchins, 2003).²⁸ U današnje vrijeme postoje i mrežno-bazirani CAT alati koji pružaju usluge i alate putem interneta. To omogućuje pristupanje projektima i resursima iz bilo kojeg mjesta i uređaja s pristupom internetu, čime se olakšava fleksibilnost i mobilnost u radu (Folaron, 2010).²⁹

CAT sustavi, kako ih definira Folaron (2010)³⁰, predstavljaju grupu kompleksnih programa koji kombiniraju različite funkcije s glavnim ciljem: činiti proces prevođenja efikasnijim i bržim. Oni omogućuju bolju učinkovitost i dosljednost pri prevođenju korištenjem različitih funkcionalnosti u koje su uključeni uređivači dokumenata, glosari, prijašnji prijevodi, itd. Navedeni sustavi pružaju podršku korisnicima tijekom cijelog procesa prevođenja: prije procesa prevođenja, tijekom procesa prevođenja i nakon procesa prevođenja (Fernández-Parra, 2010).³¹

Kod stručnih tekstova često se značajan broj pojmova pojavljuje više puta u različitim tekstovima, a to ponavljanje mogu biti određeni termini ili čak cijele rečenice. Kod ovakvog tipa teksta se značajnost CAT sustava najviše može primijetiti. Stoga, sustavi za računalno potpomognuto prevođenje integriraju dva različita skupa podataka: jedan za cijele rečenice, a drugi samo za pojmove (Fernández-Parra, 2010).³² Prijevodna memorija (engl. *Translation Memory*, TM) omogućuje ponovnu upotrebu prijevoda cijelih rečenica ili dijelova rečenica, a

²⁷ Bowker, L. & Fisher, D. (2010). Computer-aided translation. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 60-65. str. 60

²⁸ Hutchins, J. (2003). The development and Use of Machine Translation Systems and Computerbased Translation Tools. *International Journal of Translation*. 15(1). 6-26. str. 18

²⁹ Folaron, D. A. (2010). Translation tools. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 429-436. str. 436, 436

³⁰ Folaron, D. A. (2010). Translation tools. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 429-436. str. 429, 435

³¹ Fernández-Parra, M. (2010). The Workflow of Computer-Assisted Translation Tools in Specialised Translation. *Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009*. str.3

³² Fernández-Parra, M. (2010). The Workflow of Computer-Assisted Translation Tools in Specialised Translation. *Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009*. str. 4

terminološka baza (engl. *Term Base, Term Bank, TB*) potiče korištenje standardnih pojmova, naziva i prijevoda (Choudhury i McConnell, 2013).³³ Oba navedena resursa su bitna prije samog procesa prevođenja, točnije moraju biti stvorena prije kako bi se mogla koristiti u novom prevođenju. Dodatni zadaci koje CAT alati mogu obavljati prije prijevoda uključuju brojanje riječi, statistiku i analizu. Ove karakteristike mogu pružiti korisniku dodatne informacije koje on može koristiti u procesu prevođenja. Primjerice, može provjeriti koliko je podudaranje novog teksta koji se prevodi s resursima te prema tome se može odrediti opseg budućeg rada (Fernández-Parra, 2010).³⁴

Tijekom procesa prevođenja CAT sustavi pregledavaju tekst koji se prevodi i automatski daju prijevode u slučaju kada ima podudaranja s prijevodnim memorijama i terminološkim bazama. Stoga sustav automatski prikazuje dijelove ciljnog prijevoda korisniku. Ovaj postupak stvara prednost kod postojanja velike sličnosti novog teksta s postojećim resursima što je vrlo često kod stručnih tekstova. Tijekom procesa prevođenja se primjećuje prednost postupaka prije samog prevođenja, jer CAT sustavi ovise o kvalitetnim prijevodnim memorijama i terminološkim bazama (Brkić i sur., 2009).³⁵ Nakon što CAT sustav prikaže rezultate prijevoda koji se poklapa s resursima dolazi do prevođenja od strane korisnika. Korisnik provjerava prijevod dan od strane CAT alata temeljen na prijevodnim memorijama i terminološkim bazama te potom prevodi ostatak teksta. Tijekom postupka prevođenja korisnik može nadopunjavati prijevodnu memoriju i terminološku bazu. Nadalje, za što točnije prijevode, CAT sustavi imaju ugrađene alate koji osiguravaju kvalitetu prijevoda (engl. *Quality Assurance, QA*) kako bi se ispravile moguće pogreške (Fernández-Parra, 2010).³⁶

Nakon prevođenja prema Choudhury i McConnell (2013)³⁷ potrebno je odraditi procjenu kvalitete prijevoda (engl. *Translation Quality Assessment, TQA*). Uz prijevodne memorije i terminološke baze, procjena kvalitete prijevoda igra ključnu ulogu u CAT sustavima. Procjena kvalitete prijevoda se odrađuje nakon prevođenja i tek tada je korisnik u mogućnosti završiti cjelokupni proces prevođenja.

³³ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 15

³⁴ Fernández-Parra, M. (2010). The Workflow of Computer-Assisted Translation Tools in Specialised Translation. Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009. str. 6

³⁵ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. INFUTURE-2009 Digital Resources and Knowledge Sharing. 353-363. str. 355

³⁶ Fernández-Parra, M. (2010). The Workflow of Computer-Assisted Translation Tools in Specialised Translation. Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009. str. 10, 11

³⁷ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 15

3.1. Prijevodna memorija

Prema Seljan i suradnicima (2007)³⁸ prijevodna memorija predstavlja središnju komponentu CAT alata i osmišljena je kako bi tijekom procesa prevođenja bilo omogućeno ponovno korištenje prijevoda i održavanje dosljednosti tijekom cijelog procesa prevođenja. Prijevodna memorija funkcionira tako da pohranjuje prethodno prevedene segmente ili rečenice s njihovim odgovarajućim izvornim i ciljnim jezičnim parovima u prijevodne jedinice (engl. *translation units*, TU) (Hartley, 2009).³⁹ Navedeni parovi su spremljeni u datoteku u posebnom formatu koji je karakterističan za određeni CAT alat, kao što je to **sdlxliff* za *Trados Studio* (Choudhury i McConnell, 2013).⁴⁰

Postoje dva načina na koji prijevodne memorije mogu biti korištene. Prvi primjer korištenja prijevodne memorije je pretraživanje prijevodne memorije zbog određenog upita ili traženja određene prijevodne jedinice koja može biti korisna tijekom procesa prevođenja. Ovo se obavlja korištenjem alata za konkordancije. Primjerice ako korisnik nije siguran na koji način se neki termin koristi u rečenici. Drugi i glavni način korištenja se događa tijekom procesa prevođenja kada CAT sustavi pretražuju prijevodnu memoriju i uspoređuju je s tekstom koji treba prevesti (Fernández-Parra, 2010).⁴¹ Na početku procesa prevođenja CAT alat automatski pretražuje sve učitane prijevodne memorije i signalizira postoji li sličnost između novog teksta i već prevedenih segmenata. CAT sustavi mogu automatski umetnuti odgovarajuće prijevode izravno u ciljni tekst ili predložiti nekoliko sličnih, ali ne i potpuno jednakih prijevoda otvaranjem zasebnog prozora iz kojeg će korisnik odabrati prijedlog po potrebi ili ga ignorirati (Brkić i sur., 2009),⁴² što značajno štedi vrijeme (Choudhury i McConnell, 2013).⁴³ Osim toga, ova funkcionalnost također pridonosi dosljednosti u prijevodu jer se već provjereni i odobreni prijevodi koriste ponovno. Kroz integraciju prijevodne memorije s drugim značajkama CAT alata, kao što su terminološke baze i provjera kvalitete,

³⁸ Seljan, S., Gašpar, A. & Pavuna, D. (2007). Sentence Alignment as the Basis for Translation Memory Database. INFUTURE2007: "Digital Information and Heritage". 299-311. str. 300

³⁹ Hartley, T. (2009). Technology and Translation. Munday, J. (Ur.) The Routledge Companion to Translation Studies. 106-127. str. 117

⁴⁰ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 20

⁴¹ Fernández-Parra, M. (2010). The Workflow of Computer-Assisted Translation Tools in Specialised Translation. Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009. str. 9

⁴² Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. INFUTURE-2009 Digital Resources and Knowledge Sharing. 353-363. str. 355

⁴³ Seljan, S., Gašpar, A. & Pavuna, D. (2007). Sentence Alignment as the Basis for Translation Memory Database. INFUTURE2007: "Digital Information and Heritage". 299-311. str. 301

korisnicima se omogućuje brže pronalaženje i korištenje prijevoda koji su slični trenutnom kontekstu (Choudhury i McConnell, 2013).⁴⁴

Cilj prijevodnih memorija je njihovo korištenje u budućnosti kako bi se poboljšala učinkovitost i kvaliteta procesa prevođenja pomoću CAT sustava. Stoga je ključno stvarati što veći broj prijevodnih jedinica unutar prijevodnih memorija. Što više prijevodnih jedinica postoji u memoriji, to će CAT alati imati veći i raznolikiji skup informacija koje mogu prikazati korisnicima tijekom prevođenja (Brkić Bakarić i sur., 2011).⁴⁵ Jaworski i suradnici (2023)⁴⁶ prepoznali su ovu važnost i razvili platformu za prijevodne memorije nazvanu *TMrepository*, koja trenutno sadrži 4 milijuna prijevodnih jedinica. Ova platforma je namijenjena kao resurs za korisnike kako bi im omogućila lak pristup prijevodnim jedinicama na razini rečenica. Osim kvantitete, važno je naglasiti i kvalitetu prijevodnih memorija. Veće prijevodne memorije pružaju bolje rezultate u prevođenju pomoću CAT alata jer imaju veće podudaranje sa sličnim prevođenjima u budućnosti. Stoga, kako se prijevodne memorije povećavaju, tako se povećava i vjerojatnost da će novi prijevodi biti točniji i precizniji. Nadalje, važno je istaknuti da su veće prijevodne memorije poput prijevodne memorije *Europarl* i *Acquis Communautaire* dokazale svoju ključnu ulogu u različitim jezičnim aplikacijama. Ove velike prijevodne memorije koriste se i za pretraživanje informacija te za stvaranje jezičnih resursa kao što su jezični korpusi.

3.1.1. Segmentacija

Segmentacija (engl. *segmentation*) se odnosi na podjelu teksta na manje jedinice, koje se nazivaju segmenti. Korištenje alata za segmentaciju je nužno za proces integracije prijevoda u prijevodne memorije jer se segmentirani parovi izvornog i ciljnog jezika spremaju u prijevodne jedinice (Han, 2020).⁴⁷ Segmenti mogu obuhvaćati nekoliko riječi, cijelu rečenicu ili čak više

⁴⁴ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 21

⁴⁵ Brkić Bakarić, M., Matetić, M. & Seljan, S. (2011). Towards Obtaining High Quality Sentence-Aligned English-Croatian Parallel Corpus. Conference: 4th IEEE International Conference on Computer Science and Information Technology ICCSIT 2011. At: Chengdu, China. 241-248. str. 241

⁴⁶ Jaworski, R., Seljan, S., & Dunder, I. (2023). Four Million Segments and Counting: Building an English-Croatian Parallel Corpus through Crowdsourcing Using a Novel Gamification-Based Platform. *Information*. 14 (4). 226. str. 14,15

⁴⁷Han, B. (2020). Translation, from Pen-and-Paper to Computer-Assisted Tools (CAT Tools) and Machine Translation (MT). *Proceedings 2020*, 63(1), 56. str. 5

rečenica ovisno o tome koristi li se pristup temeljen na rečenicama ili pristup temeljen na znakovnom nizu u bitekstu (engl. *Character-string-in-bitext-based approach*, CSB). Iako se razlikuju u načinu implementacije, oba pristupa imaju zajedničku svrhu: pohranjivanje prethodno prevedenog materijala na organiziran način kako bi se koristio u budućim prijevodima (Brkić i sur., 2009).⁴⁸

Pristup koji se temelji na rečenicama podrazumijeva podjelu tekstova izvornog i ciljnog jezika koje odgovaraju rečenicama, naslovima, podnaslovima ili unosima na popisu (Brkić i sur., 2009).⁴⁹ Početno određivanje segmenata uglavnom se oslanja na interpunkcijske znakove, međutim u različitim jezicima nisu ista pravila za interpunkcijske znakove te se to mora uzeti u obzir (Hartley, 2009).⁵⁰ Brkić i suradnici (2009)⁵¹ naglašavaju kako je prepoznavanje rečeničnih jedinica jednostavno ako počinju velikim slovom i završavaju točkom. Međutim, kratice ili točke koje se ne nalaze na kraju rečenice mogu predstavljati izazov. Kada interpunkcija nije dovoljna za potpuno određivanje segmenata, alat ima sposobnost da na temelju strukture dokumenta te uključivanjem elemenata kao što su imena, datumi ili brojevi, izvede pretpostavke o granicama rečenica. Također, ovi segmenti mogu sadržavati pojmove koji se podudaraju s unosima u terminološkoj bazi, pružajući dodatnu podršku alatu pri procesu segmentacije. U slučaju kada segmentacija i dalje nije u potpunosti točno odrađena, korisnik ima mogućnost ručno spajati ili razdvajati segmente (Hartley, 2009).⁵² Glavna prednost pristupa temeljenog na rečenicama, u usporedbi s pristupom temeljenim na nizu znakova u bitekstu, jest veća vjerojatnost da će potpuna podudaranja s prijevodnom memorijom biti relevantna (Brkić i sur., 2009).⁵³

Pristup temeljen na nizu znakova u bitekstu temelji se na prepoznavanju i ponovnoj upotrebi identičnih nizova znakova bilo koje duljine. Ovaj pristup omogućuje spremanje cijelih tekstova i njihovih prijevoda u prijevodne memorije te kasniju analizu i usklađivanje pojedinačnih znakovnih fragmenata umjesto rečenica. Korištenje znakovnih fragmenata

⁴⁸ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. *INFuture-2009 Digital Resources and Knowledge Sharing*. 353-363. str. 354

⁴⁹ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. *INFuture-2009 Digital Resources and Knowledge Sharing*. 353-363. str. 355

⁵⁰ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 112

⁵¹ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. *INFuture-2009 Digital Resources and Knowledge Sharing*. 353-363. str. 355

⁵² Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 112

⁵³ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. *INFuture-2009 Digital Resources and Knowledge Sharing*. 353-363. str. 355

olakšava identifikaciju identičnih dijelova rečenica ili čak uzastopnih identičnih rečenica, što značajno pojednostavljuje proces. Važno je napomenuti da ovaj pristup može imati svoje ograničenje u slučajevima kada se duži niz znakova nalazi unutar prijevodne memorije, što može rezultirati neprepoznavanjem kraćih znakovnih fragmenata tijekom novog prevođenja (Brkić i sur., 2009).⁵⁴

U oba pristupa potrebno je prvo pripremiti tekst za proces segmentacije. Priprema izvornog teksta kako bi se postigla bolja segmentacija može obuhvaćati nekoliko koraka. Jedan od koraka je pravopisna i gramatička provjera teksta. Nadalje, meki prijelazi, crte i određena interpunkcija mogu se ukloniti kako bi se poboljšala segmentacija teksta. Ova priprema može ovisiti o samom tekstu i o potrebama prijevoda (Seljan i sur., 2007).⁵⁵ Nadalje, mogu se definirati nova pravila za razgraničenje rečenica u samom alatu koji se koristi. Ta pravila mogu uključivati dodatne uvjete koji pomažu u prepoznavanju kraćih segmenata i izbjegavanju pogrešnih podjela. Na primjer, može se definirati pravilo da se točka nakon kratica ne smatra znakom kraja rečenice, osim ako se kratica ne pojavljuje na kraju segmenata s jedinstvenim identifikatorom (Brkić i sur., 2009).⁵⁶

3.1.2. Sravnjivanje

Prema Fernández-Parra (2010)⁵⁷ sravnjivanje predstavlja osnovni dio procesa stvaranja prijevodnih memorija iz postojećih prijevoda. Proces sravnjivanja započinje s dva paralelna korpusa, jedan na izvornom jeziku, a drugi na ciljnom jeziku. Paralelni korpusi predstavljaju temeljni resurs za daljnji rad u sustavima za računalno potpomognuto prevođenje (Jaworski i sur., 2023).⁵⁸ Ključne značajke kvalitetnog korpusa uključuju računalnu čitljivost te značajnu količinu znakova. U kontekstu upotrebe korpusa u prijevodnim memorijama, nužno je osigurati

⁵⁴ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. INFUTURE-2009 Digital Resources and Knowledge Sharing. 353-363. str. 355

⁵⁵ Seljan, S., Gašpar, A. & Pavuna, D. (2007). Sentence Alignment as the Basis for Translation Memory Database. INFUTURE2007: "Digital Information and Heritage". 299-311. str. 301, 302

⁵⁶ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. INFUTURE-2009 Digital Resources and Knowledge Sharing. 353-363. str. 355

⁵⁷ Fernández-Parra, M. (2010). The Workflow of Computer-Assisted Translation Tools in Specialised Translation. Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009. str. 6

⁵⁸ Jaworski, R., Seljan, S., & Dunđer, I. (2023). Four Million Segments and Counting: Building an English-Croatian Parallel Corpus through Crowdsourcing Using a Novel Gamification-Based Platform. Information. 14 (4). 226. str. 1

da korpus bude tematski usmjeren i specifičan za određenu domenu kao što je medicina. Ovakav korpus bi obuhvatio različite tekstualne izvore relevantne za domenu medicine koju uključuju različita područja iz navedene domene (Hartley, 2009).⁵⁹ Nakon pripreme korpusa, sljedeći korak procesa izrade prijevodnih memorija je segmentacija korpusa, točnije podjela izvornog i ciljnog korpusa na segmente (Jaworski i sur., 2023).⁶⁰ Potom slijedi sravnjivanje, točnije povezivanje svakog segmenta izvornog teksta s odgovarajućim segmentom ciljnog jezika stvarajući prijevodne jedinice (Fernández-Parra, 2010).⁶¹

Sustavi za računalno potpomognuto prevođenje automatski sravnjuju rečenice paralelnih tekstova. No, automatsko sravnjivanje rečenica susreće se s razlikama između izvornog i ciljnog teksta u rasporedu i sadržaju, što može utjecati na preciznost sravnjivanja. Dio ovih razlika odnosi se na različite rasporede teksta, izostanke, inverzije, različit redoslijed struktura i označavanje paragrafa. Stoga je potrebno, prije procesa sravnjivanja postaviti postavke programa ovisno o potrebama teksta. Nadalje potrebno je i prethodno uređivanje izvornog teksta kako bi se postigla bolja segmentacija te time poboljšao proces sravnjivanja. Stoga bi se nepravilno sravnjivanje smanjilo, što bi rezultiralo povećanim brojem visokokvalitetnih prijevodnih jedinica (Seljan i sur., 2007).⁶²

Automatsko sravnjivanje jedne izvorne rečenice s jednom ciljnom rečenicom može biti kompleksno, pa se često javljaju promjene. Promjene se pojavljuju kod dugih rečenica koje mogu biti razdvajane ili kod kraćih rečenica koje se mogu spajati. Stoga je često potrebna intervencija korisnika kako bi se postigli točni podudarni parovi (Seljan i sur., 2007).⁶³ Prema istraživanju provedenom od strane Gale i Churcha (1993)⁶⁴, duljina rečenica može biti snažan pokazatelj za sravnjivanje. Nadalje, vrsta sravnjivanja ovisi o broju sravnjenih rečenica, pri čemu se najčešće jedna rečenica iz izvornog korpusa sravnjuje s jednom rečenicom iz ciljnog korpusa. Doduše, ako dođe do promjena postoji mogućnost spajanja ili razdvajanja rečenica te tada neće biti moguće 1:1 sravnjivanje, točnije neće biti moguće da jedna rečenica iz izvornog

⁵⁹ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 109,110

⁶⁰ Jaworski, R., Seljan, S., & Dunđer, I. (2023). Four Million Segments and Counting: Building an English-Croatian Parallel Corpus through Crowdsourcing Using a Novel Gamification-Based Platform. *Information*. 14 (4). 226. str. 2

⁶¹ Fernández-Parra, M. (2010). *The Workfl ow of Computer-Assisted Translation Tools in Specialised Translation. Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009*. str. 6

⁶² Seljan, S., Gašpar, A. & Pavuna, D. (2007). Sentence Alignment as the Basis for Translation Memory Database. *INFuture2007: "Digital Information and Heritage"*. 299-311. str. 306, 307

⁶³ Seljan, S., Gašpar, A. & Pavuna, D. (2007). Sentence Alignment as the Basis for Translation Memory Database. *INFuture2007: "Digital Information and Heritage"*. 299-311. str. 306, 307

⁶⁴ Gale, W. & Church, K. (1993). A program for aligning sentences in bilingual corpora. *Computational linguistics*. 19(1). 75-102. str. 89

korpusa bude spojena s jednom rečenicom iz ciljnog korpusa (Brkić Bakarić i sur., 2011).⁶⁵ Prema Brkić i suradnicima (2011) postoje tri pristupa sravnjivanja: Gale i Church pristup, pristup od strane Moorea i pristup od strane Brauna i Fradera.

Gale i Church (1993)⁶⁶ pristup temelji se na pretpostavci da dulje rečenice u jednom jeziku obično imaju dulje prijevode u drugom jeziku, dok isto vrijedi i za kraće rečenice. Inicijalno se sravnjuju cijeli odlomci, a zatim se sravnjuju rečenice unutar tih odlomaka. Rezultati istraživanja pokazuju značajnu korelaciju između duljine odlomaka u znakovima i duljine njihovih prijevoda, što ukazuje na to da duljina može biti snažan indikator za sravnjivanje rečenica. Također, istraživanje je potvrdilo visoku stopu preciznosti neovisno o jeziku. Nužno je istaknuti da se sve rečenice moraju uzeti u obzir i svaka rečenica može se pojaviti samo u jednom paru rečenica, kako bi omjer bio 1:1.

Moore (2002)⁶⁷ primjenjuje kombinirani pristup koji uključuje i riječi i duljinu rečenica. Njegovo istraživanje objašnjava kako koristiti te modele za bolju preciznost uz minimalno dodatno računanje i bez potrebe za dodatnim znanjem osim segmentacije na riječi i rečenice. Proces se sastoji od tri koraka. Prvi korak uključuje sravnjivanje korpusa korištenjem prilagođene metode temeljene na duljini rečenica, slične pristupu Galea i Churcha. Zatim se koriste parovi rečenica s najvišom vjerojatnošću usklađivanja za treniranje prilagođene verzije *IBM Modela 1* za prijevod. Konačno, korpus se ponovno sravnjuje, a početni model se nadopunjuje s *IBM Modelom 1*, rezultirajući sravnjivanjem koje obuhvaća i duljinu rečenica i usklađivanje riječi. Završna faza usmjerena je na minimalne segmente sravnjivanja, koje stvara brže završno sravnjivanje u odnosu na početno.

Braune i Fraser pristup je modificirana verzija Mooreova pristupa. Za razliku od Mooreovog pristupa, njihov pristup omogućuje sravnjivanje u situacijama kada jedna rečenica odgovara više rečenica ili obrnuto (Brkić i sur., 2011).⁶⁸ Njihova metoda sastoji se od dva koraka za proces sravnjivanja rečenica, koji unapređuju model temeljen na duljini rečenica uz pomoć leksičke statistike. Prvi korak koristi statistiku temeljenu na duljini rečenica za

⁶⁵ Brkić Bakarić, M., Matetić, M. & Seljan, S. (2011). Towards Obtaining High Quality Sentence-Aligned English-Croatian Parallel Corpus. Conference: 4th IEEE International Conference on Computer Science and Information Technology ICCSIT 2011. At: Chengdu, China. 241-248. str. 242

⁶⁶ Gale, W. & Church, K. (1993). A program for aligning sentences in bilingual corpora. *Computational linguistics*. 19(1). 75-102. str. 89

⁶⁷ Moore, R. (2002). Fast and accurate sentence alignment of bilingual corpora. *Proc. 5th Conf. Association for Machine Translation in the Americas (AMTA 02)*. 135-144. str. 143

⁶⁸ Brkić Bakarić, M., Matetić, M. & Seljan, S. (2011). Towards Obtaining High Quality Sentence-Aligned English-Croatian Parallel Corpus. Conference: 4th IEEE International Conference on Computer Science and Information Technology ICCSIT 2011. At: Chengdu, China. 241-248. str. 243

prikupljanje podataka za obuku *IBM Modela 1*. U sljedećem koraku provodi se sravnjivanje koje se fokusira na minimalne podudarnosti, kao što su 1:0, 0:1 i 1:1. Ta sravnjivanja se potom kombiniraju kako bi se stvorila sveobuhvatnija sravnjivanja. Njihovo istraživanje potvrđuje učinkovitost ovog pristupa u situacijama kada jedna rečenica odgovara više rečenica ili obrnuto (Braune i Fraser, 2010).⁶⁹

3.1.3. Podudaranje

Jedan od načina stvaranja prijevodne memorije je procesom sravnjivanja, a drugi način je procesom prevođenja. Kod stvaranja novih prijevoda, koristeći postojeće prijevodne memorije, segmenti mogu biti podudarni sa segmentima iz prijevodne memorije (Choudhury i McConnell, 2013).⁷⁰ Trados (bez dat.)⁷¹ navodi 4 podjele podudaranja:

1. Potpuno podudaranje (engl. *full match*)
2. Nejasno podudaranje (engl. *fuzzy match*)
3. Nema podudaranja (engl. *no match*)
4. Kontekstualno podudaranje (engl. *context match*)

Potpuno podudaranje događa se kada je segment u izvornom tekstu identičan prethodno prevedenom segmentu u prijevodnoj memoriji, točnije segmenti su 100% podudarni. U tom slučaju CAT alat automatski dodjeljuje prijevod iz prijevodne memorije. Taj prijevod je automatski potvrđen, ali se predlaže pregled od strane korisnika. Ova vrsta podudaranja vrlo je korisna jer ne zahtijeva dodatni napor i može uvelike skratiti vrijeme prevođenja (Moran i sur., 2014).⁷² Nejasno podudaranje se događa kada se segment u izvornom tekstu djelomično podudara s prethodno prevedenim segmentom u prijevodnoj memoriji. U slučaju kada je

⁶⁹ Braune, F. & Fraser, A. (2010). Improved unsupervised sentence alignment for symmetrical and asymmetrical parallel corpora. Proc. 23rd International Conf. Computational Linguistics (COLING 2010). 81-89. str. 88

⁷⁰ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 21

⁷¹ Trados (bez dat.). What is Translation Alignment?. Dostupno 20.5.2023. na <https://www.trados.com/solutions/translation-alignment/>.

⁷² Moran, J., Saam, C. & Lewis, D. (2014). Towards desktop-based CAT tool instrumentation. Proceedings of the 11th Conference of the Association for Machine Translation in the Americas. 99-112. str. 106

podudaranje manje od 100%, tada ono više nije potpuno (Choudhury i McConnell, 2013).⁷³ Nejasno podudaranje može biti između 1% i 99%, ali prag nejasnog podudaranja obično je postavljen na 70% (Trados, bez dat.).⁷⁴ Korisnik može pregledati predloženi prijevod te napraviti potrebne izmjene kako bi ga prilagodio trenutnim kontekstu i osigurao točnost (Moran, 2014).⁷⁵ Kada dođe do "nema podudaranja", potrebno je u potpunosti prevesti segment. U slučaju alata *Trados Studio* to dolazi u slučajevima koji su ispod 70% podudaranja. CAT alati i dalje pružaju korisne značajke kao što je prikaz termina iz terminološke baze, provjera pravopisa i druge provjere osiguravanja kvalitete kako bi se održala dosljednost i točnost prijevoda, čak i u nedostatku izravnog podudaranja s prijevodnom memorijom (Choudhury i McConnell, 2013).⁷⁶ Kontekstualno podudaranje se pojavljuje kod podudaranja segmenta i segmenata koji ih okružuju. Kontekstualno podudaranje predstavlja osnovnu komponentu u osiguravanju preciznosti prijevoda ne samo samih rečenica, već i njihovog šireg konteksta. Ova napredna vrsta podudaranja omogućuje korisnicima dublje razumijevanje značenja rečenica i njihovih veza s okolnim tekstom, omogućujući im da prilagode prijevod kako bi što bolje očuvali originalni izraz i smisao (Trados, bez dat.).⁷⁷

Kod stvaranja novog prijevoda, kada ne postoji podudarnost s prijevodnom memorijom, korisnici imaju mogućnost pohraniti nove prijevode u tu memoriju. Ovaj pristup može značajno smanjiti broj slučajeva kada nema podudaranja u budućim prevođenjima, jer korisnici postupno grade vlastitu prijevodnu memoriju koju mogu koristiti u sljedećim prevođenjima. Stoga se potencijalno može smanjiti broj slučajeva kada nema podudaranja u budućim prevođenjima (Choudhury i McConnell, 2013).⁷⁸

⁷³ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 21

⁷⁴ Trados (bez dat.). What is Translation Alignment?. Dostupno 20.5.2023. na <https://www.trados.com/solutions/translation-alignment/>.

⁷⁵ Moran, J., Saam, C. & Lewis, D. (2014). Towards desktop-based CAT tool instrumentation. Proceedings of the 11th Conference of the Association for Machine Translation in the Americas. 99-112. str. 106

⁷⁶ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 21

⁷⁷ Trados (bez dat.). What is Translation Alignment?. Dostupno 20.5.2023. na <https://www.trados.com/solutions/translation-alignment/>.

⁷⁸ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 21

3.2. Terminološke baze

Choudhury i McConnell (2013)⁷⁹ definiraju terminološke baze kao rječnike koji pohranjuju specijaliziranu terminologiju koja se koristi u projektima. One su sastavljene od riječi ili izraza uparenih s njihovim prijevodom na određeni jezik. Terminologija koja se unutra nalazi mogu biti ključne riječi, tehnički termini, vlastita imena i druge vrste ovisno o domeni u kojoj se prevodi. Dosljednost terminologije važna je za jasno razumijevanje, smanjenje neizvjesnosti i osiguravanje čitljivosti prijevoda te je to jedan od aspekata osiguravanja kvalitete prijevoda. Ova dosljednost značajno poboljšava komunikaciju, prijenos informacija i razumijevanje. U prijevodima nedostaci se prikazuju kroz različite varijacije prijevoda. Primjerice, ako postoji više istih riječi koje opisuju određeni termin te se više njih koristi u prijevodu to može biti zbunjujuće krajnjem korisniku (Gašpar i sur., 2022).⁸⁰ Upravljanje terminologijom podrazumijeva niz koraka kao što su prikupljanje podataka, izrada terminoloških baza te pregledavanje i pohrana termina s pripadajućim informacijama o tim terminima u terminološkim bazama. Nadalje, obuhvaća i druge zadatke kao što su razvrstavanje, provjera ispravnosti, ažuriranje, preoblikovanje i slični dodatni postupci (Warburton, 2006).⁸¹

3.2.1. Izrada terminoloških baza

Choudhury i McConnell (2013)⁸² navode kako se terminološke baze mogu raditi samostalno ili postoji mogućnost integriranja postojećih terminoloških baza. U slučaju kada je korisnik već prevodio slične tekstove iz određene domene gdje se koriste isti termini tada može

⁷⁹ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 20

⁸⁰ Gašpar, A., Seljan, S. & Kučiš, V. (2022.) Measuring Terminology Consistency in Translated Corpora: Implementation of the Herfindahl-Hirshman Index. *Information* 13 (2). 43. str. 2

⁸¹ Warburton, K. (2006) Terminology as a Key Driver in Business Communications, Bridging the Divide Between Authoring and Translation, *The Globalization Insider*.

⁸² Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 23

koristiti postojeću bazu termina te si tako skratiti vrijeme prevođenja. Kada terminološka baza ne postoji može se stvoriti nova baza i koristiti se u kasnijim prijevodima (Webb, 1998).⁸³

Proces izgradnje terminoloških baza podrazumijeva izvlačenje relevantnih pojmova iz jednojezičnih ili višejezičnih korpusa. Osnovna svrha izvlačenja termina prilagođenih određenoj domeni leži u njihovoj kasnijoj upotrebi unutar računalno potpomognutih sustava za prevođenje (Hartley, 2009).⁸⁴ Kako bi se izvukli termini vezani uz određenu domenu jako je važno imati veliku količinu resursa iz kojih bi se mogli izvući relevantni termini za određenu domenu. U CAT sustavima se relevantni termini izvlače iz velike količine prijevodnih jedinica u prijevodnim memorijama (Blancafort i sur., 2011).⁸⁵ Unutar okvira prijevodnih memorija, termini i njihovi prijevodi su odvojeni unutar dva korpusa: termini na izvornom jeziku i njihovi prijevodi na ciljnom jeziku. Sustav prvo treba izvući relevantne termine u izvornom korpusu, što mogu biti riječi ili kombinacija riječi te ih potom spajati s njihovim prijevodima (Blancafort i sur., 2011).⁸⁶

Kod izvlačenja terminologije primjenjuju se dvije metode koje omogućuju prepoznavanje relevantnih termina unutar teksta. Prva metoda usredotočuje se na lingvistički pristup, gdje se analiziraju oznake vrsta riječi i njihova međusobna povezanost u skladu s prethodno definiranim obrascem. Na primjer, moguće je identificirati određene obrasce kada se dvije imenice pojavljuju jedna za drugom. Ova metoda ovisi o jezičnim pravilima i karakteristikama jezika koji je korišten u resursu iz kojeg se izvlači. Druga metoda, koristi statistički pristup te provodi analizu kako bi se utvrdilo pojavljivanje određene kombinacije riječi u tekstu više puta nego što je očekivano nasumično. Ovaj pristup prepoznaje obrasce koji se često ponavljaju u određenom redosljedju ili obliku. U praksi, većina CAT sustava koristi kombinaciju oba pristupa kako bi postigla preciznost i obuhvatnost u izvlačenju terminologije (Hartley, 2009).⁸⁷ Kada sustav identificira relevantne termine na jednom jeziku, potrebno je uskladiti termin s njegovim prijevodom na drugom jeziku. Tako se omogućava pohranjivanje

⁸³ Webb, L. E. (1998). *Advantages and Disadvantages of Translation Memory: A Cost/Benefit Analysis*. Monterey Institute of International Studies. Monterey, CA. str. 13

⁸⁴ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 113

⁸⁵ Blancafort, H., Heid, U., Gornostay, T., Méchoulam, C., Daille, B. & Sharoff, S. (2011). *User-centred Views on Terminology Extraction Tools: Usage Scenarios and Integration into MT and CAT Tools*. *Tralogy I. Métiers et technologies de la traduction: quelles convergences pour l'avenir?* str. 2

⁸⁶ Blancafort, H., Heid, U., Gornostay, T., Méchoulam, C., Daille, B. & Sharoff, S. (2011). *User-centred Views on Terminology Extraction Tools: Usage Scenarios and Integration into MT and CAT Tools*. *Tralogy I. Métiers et technologies de la traduction: quelles convergences pour l'avenir?* str. 2

⁸⁷ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 113

termina i njegovog prijevoda na određenim jezicima (Blancafort i sur., 2011).⁸⁸ Nakon stvaranje terminološke baze, korisnici dodatno provjeravaju i uređuju terminološke parove kako bi osigurali točnost (Hartley, 2009).⁸⁹

3.2.2. Integracija terminoloških baza u CAT sustave

Prema Guillou (2013)⁹⁰ poticanje dosljednog prijevoda imenica, glagola i pridjeva u određenim domenama je nužno za stručne dokumente. Tradicionalno prevođenje često ne sadrži navedenu dosljednost kod stručnih dokumenata. Iako ne sadrži dosljednost sam prijevod je razumljiv, ali u stručnim dokumentima potrebno se pridržavati dosljednosti koju nije lako postići. Nasuprot tradicionalnom prevođenju, strojno prevođenje se temelji na statistici korištenja nekih riječi te time isto ne osigurava dosljednost terminologije. Stoga se oba pristupa ne smatraju zadovoljavajućim u kontekstu dosljednosti terminologije kod stručnih tekstova. U CAT sustavima uz podudaranje segmenata izvornog i ciljanog teksta, podudaranje se također može koristiti za pronalaženje terminologije u terminološkoj bazi. Hartley (2009)⁹¹ predstavlja ovo kao veliku prednost produktivnosti prevođenja. Budući da su terminološke baze integrirane s ostalim CAT alatima i resursima, program skenira izvorni tekst kako bi pronašao termine koji se nalaze u terminološkoj bazi. Nadalje on predlaže sam prijevod termina koji može biti odabran, pritom izbjegavajući ponovno upisivanje sklonu pogreškama. Nadalje, terminološke baze osiguravaju dosljednu upotrebu terminologije tijekom cijelog projekta. Pružanjem odobrenih prijevoda ili definicija za određene pojmove, održava se dosljednost u prijevodima, osiguravajući da se isti izraz dosljedno koristi u dokumentima, projektima ili kod različitih korisnika (Webb, 1998).⁹²

⁸⁸ Blancafort, H., Heid, U., Gornostay, T., Méchoulam, C., Daille, B. & Sharoff, S. (2011). User-centred Views on Terminology Extraction Tools: Usage Scenarios and Integration into MT and CAT Tools. *Tralogy I. Métiers et technologies de la traduction: quelles convergences pour l'avenir?* str. 2

⁸⁹ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 113

⁹⁰ Guillou, L. (2013). *Analysing Lexical Consistency in Translation*. In *Proceedings of the Workshop on Discourse in Machine Translation*. 10–18. str. 11

⁹¹ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 106,107

⁹² Webb, L. E. (1998). *Advantages and Disadvantages of Translation Memory: A Cost/Benefit Analysis*. Monterey Institute of International Studies. Monterey, CA. str. 13

Osim dosljednosti, prednost korištenja terminološke baze unutar CAT sustava je ušteda vremena potrebnog za prevođenje, posebno kod tekstova koji sadrže velike količine specijaliziranog sadržaja i specijaliziranih termina. Budući da korisnik ne mora trošiti puno vremena na istraživanje specijalizirane terminologije, već iskoristiti termine koji se nalaze u terminološkoj bazi, vrijeme prevođenja je znatno kraće. Umjesto traženja ispravnog prijevoda ili poteškoće s terminologijom, korisnici se mogu brzo pozvati na terminološku bazu i pronaći odgovarajući prijevod, smanjujući vrijeme istraživanja i poboljšavajući produktivnost. Također, korištenje terminoloških baza uvelike poboljšava samu točnost prijevoda i smanjuje pogrešno prevođenje (Hartley, 2009).⁹³

3.2.3. Metapodaci u terminološkim bazama

Prema istraživanju Kerremansa (2010)⁹⁴ specijalizirani rječnici se često koriste kako bi se bolje razumjeli određeni koncepti, predstavili odgovarajući termini i pronašli prikladni prijevodi za izvorni tekst. Međutim, tim rječnicima često nedostaju dodatne informacije koje mogu pomoći kod prevođenja. Primjerice informacije o tome kako se termin koristi u stvarnim kontekstima, što predstavlja problem kod prevođenja. Njegovo istraživanje je pokazalo da specijalizirani rječnici, koji prikazuju više podataka o samom terminu doprinose samom prijevodu. Na primjeru korištenja terminoloških baza u sustavima za računalno potpomognuto prevođenje, moguće je prilagoditi pregled sadržaja terminološke baze, čime se prilagođava prikaz informacija povezanih s pojmovima unutar nje (Hartley, 2009).⁹⁵

Prema Hartley (2009)⁹⁶ terminološke baze mogu sadržavati različite metapodatke vezane uz određeni termin. Ti podaci mogu biti:

1. Lingvistički (ekvivalenti, sinonimi, varijante i sl.)
2. Konceptualni (definicija, domena, povezani koncepti i sl.)

⁹³ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 114

⁹⁴ Kerremans, K. (2010). *Comparative Study of Terminological Variation in Specialised Translation*. *Reconceptualizing LSP*. XVII Eur. LSP Symp. 1–14. str 3,4

⁹⁵ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 114

⁹⁶ Hartley, T. (2009). *Technology and Translation*. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127. str. 114

3. Pragmatični (ograničenja upotrebe)
4. Bibliografski (izvor)
5. Upravljački (pouzdanost, autor, datum)

Metapodaci koji prate određeni termin mogu biti automatski generirani prilikom stvaranja terminološke baze, uključujući podatke poput autora i datuma. Međutim, nije uvijek moguće automatski stvarati metapodatke, jer ponekad nije poznat izvor specifičnog resursa iz kojeg su izvedeni termini. Unatoč tome, ističe se važnost proizvodnje što više mogućih metapodataka. Metapodaci su od izuzetne važnosti u kontekstu prijevoda stručnih tekstova ispunjenih stručnim terminima. Posebno je korisno omogućiti prilagodbu vrste metapodataka koje CAT sustav prikazuje, kako bi se optimalno podržala analiza i upotreba termina (Blancafort i sur., 2011).⁹⁷

3.3. Procjena kvalitete prijevoda

Tehnologije za prevođenje se koriste više od četrdeset godina i igraju ključnu ulogu u olakšavanju procesa prevođenja. S porastom globalizacije potražnja za prijevodima je značajno porasla. Istovremeno, često se postavljaju kratki rokovi za stvaranje prijevoda te nedostatak kvalificiranih stručnjaka predstavlja problem (Çetiner, 2021).⁹⁸ Stoga se javlja potreba za bržom i efikasnijom procjenom kvalitete prijevoda. Procjena kvalitete prijevoda je postupak kojim se ocjenjuje kvaliteta prijevoda obavljenih unutar alata kao što su CAT alati. Procjena kvalitete prijevoda koristi različite metode i kriterije kako bi se utvrdilo jesu li prijevodi točni i dosljedni prema specifičnim jezičnim i kontekstualnim zahtjevima koji ovise o samom kontekstu sadržaja koji se prevodi (Petrova, 2019).⁹⁹

⁹⁷ Blancafort, H., Heid, U., Gornostay, T., Méchoulam, C., Daille, B. & Sharoff, S. (2011). User-centred Views on Terminology Extraction Tools: Usage Scenarios and Integration into MT and CAT Tools. *Tralogy I. Métiers et technologies de la traduction: quelles convergences pour l'avenir?* str. 8

⁹⁸ Çetiner, C. (2021). Evaluating Quality Assurance (QA) Checks in CAT tools in Relation to Professional Quality Standards. *Euroasian Conference on Language and Social Sciences*. 338-349. str. 338

⁹⁹ Petrova, V. (2019). Translation Quality Assessment Tools and Processes in Relation to CAT Tools. *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*. 89-97. str. 95

U industriji prevođenja postoje mnogi međunarodni standardi koji definiraju kvalitetu prijevoda i identificiraju faktore koji na nju mogu utjecati. Takvi faktori uključuju kvalitetu konačnog proizvoda, vještine koje osoba koja prevodi mora posjedovati, kvalitetu postupka završne revizije, kvalitetu procesa odabira stručnjaka te upravljanje cjelokupnim procesom prevođenja. Nadalje, standardi opisuju kriterije za procjenu kvalitete prijevoda te specifikacije usluga koje pružatelji prijevodnih usluga trebaju zadovoljiti. Standardi naglašavaju da pružatelji jezičnih usluga trebaju koristiti alate za prijevodnu tehnologiju i jezične resurse koji su relevantni za domenu za koju se prevodi. U okviru CAT sustava, ovo obuhvaća prijevodne memorije i terminološke baze, kao i pružanje podrške za prevođenje prije, za vrijeme i nakon procesa prevođenja (Çetiner, 2021).¹⁰⁰

3.3.1. Osiguravanje kvalitete

Korištenje sustava za računalno potpomognuto prevođenje može rezultirati s porastom pogrešaka tijekom procesa prevođenja kod pojave većeg opsega posla, kratkih rokova ili zajedničkog rada na istom projektu. Provjera tih pogrešaka od strane ljudi je dugotrajna, sklona subjektivnosti i mogućim pogreškama. Zato je važno koristiti alate za osiguravanje kvalitete (engl. *Quality Assurance*, QA) kako bi se to izbjeglo (Seljan i sur., 2021).¹⁰¹ Choudhury i McConnell (2013)¹⁰² navode kako se osiguravanje kvalitete prijevoda oslanja na kombinaciju tehnologije i procesa kako bi se spriječile pogreške u projektima. Kvaliteta prijevoda ne može se postići samo korištenjem alata za računalno potpomognuto prevođenje i alata za osiguravanje kvalitete. Ti alati mogu pružiti određenu pomoć, ali ne mogu zamijeniti čovjeka. Ipak, kvaliteta prijevoda može se poboljšati ako se alati pravilno koriste (Czopik, 2014).¹⁰³

¹⁰⁰ Çetiner, C. (2021). Evaluating Quality Assurance (QA) Checks in CAT tools in Relation to Professional Quality Standards. Euroasian Conference on Language and Social Sciences. 338-349. str. 339, 340

¹⁰¹ Seljan, S., Škof Erdelija, N., Kučič, V., Dunder, I. & Pejić Bach, M. (2021). Quality Assurance in Computer-Assisted Translation in Business Environments. Natural Language Processing for Global and Local Business. 242-270. str. 245

¹⁰² Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 24

¹⁰³ Czopik, J. (2014). Quality Assurance process in translation. Translating and Computer. 36. 77-85. str. 79

3.3.1.1. Osiguravanje kvalitete prije procesa prevođenja

Proces osiguravanja kvalitete počinje prije nego što se projekt pošalje na prijevod. Primjeri osiguravanja kvalitete prije prevođenja su pročišćavanje teksta radi zaštite neprevodivih elemenata, razjašnjenje izvornog teksta, davanje komentara i konteksta (Choudhury i McConnell, 2013).¹⁰⁴ Brojne pogreške mogu biti povezane s problematičnim izvornim tekstom. Izvorni tekst može biti slabo razumljiv, što rezultira poteškoćama u prijevodu. Prevođenje nerazumljivog izvornog teksta u razumljiv ciljni tekst stvara probleme kod prevođenja (Czopik, 2014).¹⁰⁵ Nadalje, potrebno je odlučiti koji stručnjaci najbolje odgovaraju zadatku, uzimajući u obzir razinu vještine, prethodne rezultate osiguravanja kvalitete, dostupnost i domenu stručnosti (Choudhury i McConnell, 2013).¹⁰⁶ Bowker i Fisher (2010)¹⁰⁷ navode kako kvaliteta prevođenja u CAT alatima još uvijek ovisi o korisnicima. Primjerice, korisnici mogu doći u iskušenje da ostanu bliski strukturi izvornog teksta ili pojedinačni segmenti mogu nositi različite stilove svojih autora, što dovodi do loše čitljivosti ciljnog teksta (Bowker i Fisher, 2010).¹⁰⁸ Osim stručnjaka koji prevodi, sam proces prevođenja ovisi o različitim stručnjacima sa specijaliziranim vještinama za različite zadatke. Važno je osigurati aktivno sudjelovanje svih članova tima, efikasno organizirati radni proces, jasno raspodijeliti zadatke i pružiti odgovarajuću obuku kako bi se produbilo njihovo znanje (Seljan, 2018).¹⁰⁹

¹⁰⁴ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 24

¹⁰⁵ Czopik, J. (2014). Quality Assurance process in translation. *Translating and Computer*. 36. 77-85. str. 81

¹⁰⁶ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 24

¹⁰⁷ Bowker, L. & Fisher, D. (2010). Computer-aided translation. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 60-65. str. 60

¹⁰⁸ Bowker, L. & Fisher, D. (2010). Computer-aided translation. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 60-65. str. 63

¹⁰⁹ Seljan, S. (2018). Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. European Parliament: Translation Services in the Digital World. 92-145. str. 100

3.3.1.2. Osiguravanje kvalitete tijekom procesa prevođenja

Brkić i suradnici (2009)¹¹⁰ navode kako CAT alati korisno utječu na osiguravanje kvalitete prijevoda. Danas mnogi CAT alati sadrže ugrađene module za osiguravanje kvalitete, iako se mogu koristiti vanjski programi (Czopik, 2014)¹¹¹, kao što je *Verifica*. Alati za osiguravanje kvalitete pomažu u detekciji pogrešaka, kategorizaciji i kvantifikaciji, značajno poboljšavajući analizu pogrešaka (Seljan i sur, 2021).¹¹² Kvaliteta prijevoda bi trebala obuhvaćati, ali nije ograničena na: točno prenošenje značenja s izvornog jezika na ciljni jezik, bez pravopisnih i gramatičkih pogrešaka, jezičnu izvrsnost i lakoću čitanja rečenica (Czopik, 2014).¹¹³

Prema Seljan (2018)¹¹⁴ provjere koje rade sustavi za osiguravanje kvalitete mogu se podijeliti na:

1. provjeru na razini segmenta
2. provjeru dosljednosti
3. provjeru dosljednosti terminologije
4. provjeru interpunkcije
5. provjeru gramatike i pravopisnih pogreška
6. provjeru datuma brojeva i valute
7. upravljanje terminologijom i provjeru oznaka

Provjera na razini segmenata se odnosi na otkrivanje nepostojećih ili nepotpunih prijevoda, te promjene u broju rečenica (Seljan, 2018).¹¹⁵ Kako bi se pronašle navedene pogreške, uspoređuje se duljina izvornih segmenata s duljinom odgovarajućih segmenata na ciljnom jeziku. Ako se primijeti značajna razlika u duljini, tada se taj dio označava. Ova opcija

¹¹⁰ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. INFuture-2009 Digital Resources and Knowledge Sharing. 353-363. str. 354

¹¹¹ Czopik, J. (2014). Quality Assurance process in translation. *Translating and Computer*. 36. 77-85. str. 81

¹¹² Seljan, S., Škof Erdelija, N., Kučič, V., Dunder, I. & Pejić Bach, M. (2021). Quality Assurance in Computer-Assisted Translation in Business Environments. *Natural Language Processing for Global and Local Business*. 242-270. str. 245

¹¹³ Czopik, J. (2014). Quality Assurance process in translation. *Translating and Computer*. 36. 77-85. str. 78,79

¹¹⁴ Seljan, S. (2018). Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. European Parliament: Translation Services in the Digital World. 92-145. str. 99,100

¹¹⁵ Seljan, S. (2018). Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. European Parliament: Translation Services in the Digital World. 92-145. str. 99,100

obično zahtijeva postavljanje postavki ovisno o jeziku koji se koristi jer neki jezici zahtijevaju više znakova od onog s kojeg se prevodi. Također, neki alati za osiguravanje kvalitete mogu otkriti neprevedene dijelove teksta putem prepoznavanja jezika ili analize statusa dijela teksta koji nije preveden (Kotwicki, 2018).¹¹⁶

Provjera dosljednosti se odnosi na dosljednost u korištenju segmenata što ovisi o prijevodnoj memoriji (Seljan, 2018).¹¹⁷ Koristeći prijevodne memorije, izbjegava se korištenje reference na druge segmente kako bi segmenti bili općenitiji. Primjerice, ako postoji prijašnja prijevodna memorija sam prijevod će biti podređen njoj (Bowker i Fisher, 2010).¹¹⁸ Nadalje se provjerava jesu li segmenti u izvornom tekstu uvijek prevedeni na isti način u ciljni jezik te se u ovu provjeru uključuju zaboravljeni ili prazni prijevodi (Seljan, 2018).¹¹⁹

Dosljednost u terminologiji uključuje prepoznavanje mogućeg krivog korištenja termina ili terminologije koja ne odgovara terminološkoj bazi. Terminološka baza služi kao kontrolni mehanizam u procesu osiguravanja kvalitete i omogućava upotrebu terminologije usklađene sa zahtjevima krajnjih korisnika. Stoga, visokokvalitetna i ažurirana terminološka baza treba biti integrirana u CAT sustav i sastojati se od što većeg broja termina (Seljan i sur., 2021).¹²⁰ Kvaliteta prijevoda djelomično ovisi o kvaliteti terminologije koja se nalazi u terminološkim bazama. U kontekstu terminoloških resursa promjena podataka vezana uz terminologiju mora biti potpuno transparentna i moguća za praćenje (Kotwicki, 2018)¹²¹, bez obzira radi li se o dodavanju komentara uz unos, odobrenju termina, ispravljanju pravopisa pojma ili sličnim radnjama (Chiocchetti i sur., 2017).¹²² Alati za osiguravanje kvalitete trebaju otkriti ove vrste pogrešaka i obavijestiti korisnika te tako spriječiti pogreške u prijevodnoj memoriji i konačnom prijevodu. Kod stvaranja novog prijevoda izvorni tekst se automatski uspoređuje s terminološkom bazom i termini koji se nalaze u njoj se ističu u izvornom tekstu. Za navedene

¹¹⁶ Kotwicki, M. (2018). Outsmarting yourself. Self-revision and self-impruvment. TransELTE 2018. 61-72. str. 67

¹¹⁷ Seljan, S. (2018). Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. European Parliament: Translation Services in the Digital World. 92-145. str. 100

¹¹⁸ Bowker, L. & Fisher, D. (2010). Computer-aided translation. Y. Gambier, & L. van Doorslaer (Ur.), Handbook of Translation Studies. 1. 60-65. str. 61

¹¹⁹ Seljan, S. (2018). Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. European Parliament: Translation Services in the Digital World. 92-145. str. 100

¹²⁰ Seljan, S., Škof Erdelija, N., Kučič, V., Dunder, I. & Pejić Bach, M. (2021). Quality Assurance in Computer-Assisted Translation in Business Environments. Natural Language Processing for Global and Local Business. 242-270. str. 269

¹²¹ Kotwicki, M. (2018). Outsmarting yourself. Self-revision and self-impruvment. TransELTE 2018. 61-72. str. 67

¹²² Chiocchetti, E., Wissik T., Lušický, V. & Wetzel, M. (2017). Quality assurance in multilingual legal terminological databases. The Journal of Specialised Translation. 27, 164-188. str. 165

termine sustav automatski daje prijevod termina u ciljanom jeziku. Nakon prevođenja izvorni i ciljni tekst se uspoređuju i sustav provjerava prijevod za svaki termin iz izvornog teksta koji postoji u terminološkoj bazi. Potom sustav potvrđuje ispravno korištenje terminologije i njegovu dosljednost ili prikaže moguće pogreške (Kotwicki, 2018).¹²³

Tijekom provjere interpunkcije alat obavještava korisnike u slučajevima dvostrukih razmaka, dvostrukih interpunkcijskih znakova, neusklađene uporabe zagrada i navodnika, oblikovanja koja ne odgovaraju pravilima ciljnog jezika te razmaka ispred i iza postotnih ili stupanjskih znakova. Provjera gramatike i pravopisnih pogrešaka uključuje detekciju gramatičkih i pravopisnih pogrešaka, nepravilnosti u velikim i malim slovima u izvornom i ciljnom tekstu te otkrivanje nepotrebnog ponavljanja riječi. Nadalje, provjeru datuma, brojeva i valute predstavlja provjera formata, razdjelnika, pretvorbe mjernih jedinica i drugih elemenata (Seljan, 2018).¹²⁴ Većina ovih pogrešaka alat prepoznaje ovisno o dobro postavljenim postavkama unutar CAT sustava te obavještava korisnika o njima tijekom prevođenja. Lociranje i sprečavanje pogrešaka može se efikasno postići kada su definirana pravila za gramatiku, interpunkciju i pravopis (Kotwicki, 2018).¹²⁵ Provjera oznaka se odnosi na osiguravanje jednakih oznaka i pravilnog redoslijeda označenih formata kao što su HTML, XML, MIF i drugi (Seljan, 2018).¹²⁶

Prema istraživanju Seljan i suradnika (2021)¹²⁷ alati za osiguravanje kvalitete omogućavaju otkrivanje pogrešaka, ali djeluju prema različitim načelima ovisno o korištenom alatu. Iako alati otkrivaju pogreške, oni ih ne ispravljaju. Provjera navedenih pogrešaka i potencijalni ispravak je posao korisnika.

¹²³ Kotwicki, M. (2018). Outsmarting yourself. Self-revision and self-impruvment. TransELTE 2018. 61-72. str. 67

¹²⁴ Seljan, S. (2018). Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. European Parliament: Translation Services in the Digital World. 92-145. str. 100

¹²⁵ Kotwicki, M. (2018). Outsmarting yourself. Self-revision and self-impruvment. TransELTE 2018. 61-72. str. 68

¹²⁶ Seljan, S. (2018). Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. European Parliament: Translation Services in the Digital World. 92-145. str. 8

¹²⁷ Seljan, S., Škof Erdelija, N., Kučič, V., Dunder, I. & Pejić Bach, M. (2021). Quality Assurance in Computer-Assisted Translation in Business Environments. Natural Language Processing for Global and Local Business. 242-270.str. 251

3.3.1.3. Osiguravanje kvalitete nakon procesa prevođenja

Nakon procesa prevođenja predlažu se dodatne analize. Preporučuje se praćenje dodatnih informacija poput vremena potrebnog za prijevod, vremena potrebnog za pregled, broja pregleda, truda uloženog u naknadno uređivanje, broja uključenih stručnjaka te njihove dosljednosti. Korištenje alata za osiguravanje kvalitete koji omogućuju odabir specifičnih kategorija pogrešaka je korisno jer omogućuje procjenu podataka o vrstama pogrešaka tijekom upotrebe CAT sustava. Iako različiti alati za osiguravanje kvalitete nude različite opcije vrsta pogrešaka, oni pružaju opciju zanemarivanja detektiranih pogrešaka u slučaju kada alata prikaže pogrešku koja zapravo nije stvarna. Nadalje, različiti alati na različite načine broje pogreške: neki istu pogrešku broje dva puta, dok je drugi broje samo jednom. Stoga svakako treba proći kroz pogreške i ispraviti po potrebi (Seljan i sur., 2021).¹²⁸

3.3.2. Sustavi za upravljanje prevođenjem

Sustavi za upravljanje prevođenjem pružaju naprednu razinu kontrole i automatizacije procesa. Njihova svrha je dodijeliti zadatke prevođenja različitim sudionicima te uvesti i izvesti dokumente i njihove prijevode iz drugih sustava (Choudhury i McConnell, 2013).¹²⁹ Sustavi za upravljanje prijevodom pomažu korisnicima u upravljanju resursima koji trebaju biti prevedeni. Nadalje, kontroliraju na koje jezike treba prevesti svaki projekt ili resurs. Nakon što je ta odluka donesena, dodjeljuju se prevoditelji, urednici i recenzenti. Za postizanje uspješnog prijevoda, definira se radni tijek i određuju se potrebni rokovi. Također se koriste dodatni alati ili programski dodaci, poput automatskog strojnog prevođenja, koji podržavaju proces prevođenja. Dodatno, kako bi se olakšao postupak prevođenja, koriste se resursi poput prijevodnih memorija i terminoloških baza.

¹²⁸ Seljan, S., Škof Erdelija, N., Kučić, V., Dunder, I. & Pejić Bach, M. (2021). Quality Assurance in Computer-Assisted Translation in Business Environments. *Natural Language Processing for Global and Local Business*. 242-270. str. 247

¹²⁹ Choudhury, R., & McConnell, B. (2013). *Translation Technology Landscape Report*. TAUS – Translation Automation User Society. Nizozemska. str. 26

Nakon završetka prevođenja, sustavi za upravljanje prijevodom omogućuju upravljanje povratnim informacijama o kvaliteti prijevoda te automatsko usmjeravanje pogrešaka na odgovarajuće osobe radi ispravka. Konačno, ti sustavi omogućuju upravljanje izvozom završenih prijevoda u vanjske sustave, kao što su sustavi za upravljanje sadržajem. Navedene funkcionalnosti omogućuju učinkovito upravljanje procesom prevođenja i poboljšavaju kvalitetu prijevoda, olakšavajući rad prevoditeljima i ostalim sudionicima (Choudhury i McConnell, 2013).¹³⁰

3.4. Integracija strojnog prevođenja u CAT sustave

Integracija strojnog prevođenja u CAT alate, kako navode Choudhury i McConnell (2013)¹³¹, donosi brojne prednosti korisnicima u njihovom radnom tijeku prevođenja. Kombinacija ljudskog korisnika i strojnog prevođenja omogućuje brže i učinkovitije prevođenje. CAT alati, kao što je *Trados Studio*, mogu se prilagoditi da se povežu s vanjskim sustavima za strojno prevođenje kao što je Google Prevoditelj (engl. *Google Translate*). Zaretskaya i suradnici (2015)¹³² predstavljaju dva načina kombiniranja CAT alata sa strojnim prevođenjem:

1. Uključivanje prijedloga iz strojnog prevođenja s drugim prijedlozima koje korisnik dobiva za svaki segment
2. Kombiniranje tehnologija prijevodne memorije i strojnog prevođenja

U prvom primjeru, kada se koriste prijedlozi iz strojnog prevođenja u CAT alatima, obično su korisni kada nema potpunog ili nejasnog podudaranja s postojećim prijevodima u memoriji. U tim slučajevima strojno prevođenje stvara automatski prijevod. Prijevod strojnog prevođenja

¹³⁰ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str.27

¹³¹ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 37-40

¹³² Zaretskaya, A., Corpas Pastor, G. & Seghiri M. (2015). Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. *Journal of Translation and Interpretation*. 8(1). 76-88. str. 77

korisnik može iskoristiti kao polaznu točku za vlastiti prijevod ili kao inspiraciju za daljnje istraživanje (Zaretskaya i sur., 2015).¹³³

U drugom primjeru, kada se kombinira prijevod iz prijevodne memorije sa strojnim prijevodom, program traži slične ili identične fragmente u izvornom segmentu iz prijevodne memorije. Nakon pronalaska tih segmenata, program izvlači njihove prijevode iz memorije i kombinira ih s rezultatima strojnog prevođenja kako bi stvorio prijevod za cijeli izvorni segment koji se prevodi (Biçici i sur, 2008).¹³⁴ Ova tehnika kombiniranja ima za cilj iskoristiti već prevedene dijelove teksta iz memorije, s ciljem ubrzanja postupka prevođenja i održavanja dosljednosti prijevoda (Choudhury i McConnell, 2013).¹³⁵ Alati za kombiniranje prijevoda iz memorije i strojnog prevođenja ne koriste napredne tehnologije strojnog prevođenja. Oni se oslanjaju na osnovno podudaranje nizova, slično kao kod prevođenja u strojnom prevođenju. Dakle, ova tehnika se smatra korisna za određene tekstove, ali kod integracije strojnog prevođenja u CAT alate je i dalje najvažnija ljudska intervencija koja osigurava točnost prijevoda te njegovu kvalitetu (Zaretskaya i sur., 2015).¹³⁶

Prednosti integracije strojnog prevođenja u CAT sustave su mnogobrojne. Kada korisnik želi ubrzati proces prevođenja i nema puno segmenata povezanih s prijevodnom memorijom, korisnik može aktivirati funkciju strojnog prevođenja unutar CAT alata. Alat zatim automatski može za sve segmente bez prijevoda u prijevodnoj memoriji prikazati prijevod korisniku iz strojnog prevođenja u stvarnom vremenu. Korisnik može procijeniti kvalitetu prijevoda, prilagoditi ga prema potrebi ili ga koristiti kao polaznu točku za daljnje uređivanje (Choudhury i McConnell, 2013).¹³⁷ Ova integracija omogućuje korisnicima da uz prijevodne memorije i terminološke baze kao alat u procesu prevođenja koriste strojno prevođenje (Zaretskaya i sur., 2015).¹³⁸ Integracija strojnog prevođenja u CAT alate ne zamjenjuje ulogu čovjeka, nego im ona pruža alat koji pomaže ubrzati rad i pruža veću fleksibilnost u izvršavanju njihovih

¹³³ Zaretskaya, A., Corpas Pastor, G. & Seghiri M. (2015). Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. *Journal of Translation and Interpretation*. 8(1). 76-88. str. 77

¹³⁴ Biçici, E. & Dymetman, M. (2008). Dynamic translation memory: Using statistical machine translation to improve translation memory fuzzy matches. In *Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science*. 4919. 454–465. str. 455

¹³⁵ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 44

¹³⁶ Zaretskaya, A., Corpas Pastor, G. & Seghiri M. (2015). Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. *Journal of Translation and Interpretation*. 8(1). 76-88. str. 78

¹³⁷ Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska. str. 44

¹³⁸ Zaretskaya, A., Corpas Pastor, G. & Seghiri M. (2015). Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. *Journal of Translation and Interpretation*. 8(1). 76-88. str. 76

zadataka. Korisnici i dalje igraju ključnu ulogu u prevođenju i time osiguravaju visoku kvalitetu i prilagođenost ciljanoj publici (Biçici i sur., 2008).¹³⁹

¹³⁹ Biçici, E. & Dymetman, M. (2008). Dynamic translation memory: Using statistical machine translation to improve translation memory fuzzy matches. In *Computational Linguistics and Intelligent Text Processing*. Lecture Notes in Computer Science. 4919. 454–465. str. 464, 465

4. Istraživanje

U istraživačkom dijelu ovoga rada prikazuju se mogućnosti i mogući nedostaci sustava za računalno potpomognuto prevođenje. Istraživanje se vrši unutar domene obrazovanja, usko povezano uz informacijske znanosti. Za izvođenje istraživanja korišten je *Trados Studio 2022*, softverski paket za prevođenje pomoću računala koji je razvio SDL (Trados, bez dat.).¹⁴⁰

Istraživanje započinje stvaranjem resursa. U sklopu istraživanja stvorena je prijevodna memorija procesom sravnjivanja korpusa i terminološka baza ručnim izvlačenjem termina iz prijevodne memorije. Osim navedenih resursa, u alat *Trados Studio* integriran je sustav za automatsko strojno prevođenje *Google Translate*. Nakon stvaranja i objedinjavanja resursa, provedeno je testiranje sustava te usporedna analiza prijevoda. Cilj pilot istraživanja je ispitati učinkovitost primjene CAT alata radi utvrđivanja raznolikosti stručne terminologije, vrsta i kategorija pogrešaka i stvaranja prijevoda sukladnog referentnom prijevodu.

U pilot istraživanju je sudjelovalo 10 studenata različitih fakulteta koji prevode tekstualnu datoteku unutar domene obrazovanja na tradicionalan način i pomoću CAT alata *Trados Studio*. Tekst koji treba prevesti sadrži 5 rečenica, tj. 108 riječi i sadrži 19 specifičnih termina iz domene obrazovanja. Dio navedenih termina se nalazi u terminološkoj bazi, njih 13, a dio se ne nalazi u navedenoj bazi. Svaki od ispitanika preveo je isti tekst najprije na tradicionalan način, a zatim primjenom CAT alata *Trados Studio*. U postupku prevođenja analizirani su sljedeći elementi: vrijeme potrebno za prevođenje, broj i kategorija pogrešaka, sličnost između dobivenih prijevoda i referentnog teksta te upotreba stručnih termina. Referentni tekst nastao je profesionalnim ljudskim prijevodom uz korištenje alata *Trados Studio*. Prevedeni tekstovi su uspoređeni s referentnim tekstom kako bi se utvrdila sličnost prevedenih prijevoda i referentnog prijevoda.

¹⁴⁰ Trados (bez dat.). Trados Studio Freelance. Dostupno 10.7. 2023. na: <https://www.trados.com/products/trados-studio/freelance/>

4.1. Resursi i alati

Proces prevođenja u alatu *Trados Studio* započinje stvaranjem novog projekta kako bi se omogućilo integriranje svih resursa i korištenje dodatnih alata tijekom procesa prevođenja. U istraživanju, novi projekt pruža mogućnost upotrebe prijevodne memorije, terminološke baze i automatskog strojnog prevođenja tijekom procesa prevođenja.

Stvaranjem novog projekta definiraju se postavke koje olakšavaju proces prevođenja. Zadaje se ime projekta te se dodjeljuju jezici za izvorni i ciljni jezik. U istraživanju se odabire hrvatski jezik kao izvorni i engleski jezik kao ciljni jezik. Posebna pažnja je posvećena odabiru jezika, kako bi se osiguralo precizno sravnjivanje jezičnih parova u sklopu ovoga istraživanja. Za hrvatski jezik postoje 2 odabira, hrvatski korišten u Hrvatskoj i hrvatski korišten u Bosni i Hercegovini, a za engleski jezik postoji čak 106 različitih odabira. U istraživanju za izvorni jezik je odabran hrvatski jezik koji se koristi u Hrvatskoj, a za ciljni jezik je odabran engleski jezik koji se koristi u Sjedinjenim Američkim Državama. Tijekom cijelog istraživanja mora se koristiti isti izvorni i ciljni jezik kako u novim prijevodima, tako i u prijevodnim memorijama i terminološkoj bazi. Nadalje, u postavkama novoga projekta odabiru se izvorne jezične datoteke za prevođenje i određuju dodatne postavke projekta kao što je dodavanje resursa i integracija alata.

4.1.1. Stvaranje prijevodne memorije

Prijevodna memorija se stvara procesom sravnjivanja dvaju paralelnih korpusa, pri čemu je moguće koristiti i druge alate za sravnjivanje (Brkić i sur., 2011¹⁴¹; Seljan i sur., 2010¹⁴²; Seljan i sur., 2008¹⁴³). U istraživanju su korištene paralelne tekstualne datoteke: na hrvatskom jeziku i na engleskom jeziku iz domene obrazovanja. Svaka datoteka se sastoji od 600 linija

¹⁴¹M Brkić, M Matetić, S Seljan. Towards Obtaining High Quality Sentence-Aligned English-Croatian Parallel Corpus. Computer Science and Information Technology ICCSIT 2011, 1068-1070

¹⁴²Seljan, S., Tadić, M., Agić, Ž., Šnajder, J., Dalbelo Bašić, B. & Osmann, V. (2010). Corpus Aligner (CorAl) Evaluation on English-Croatian Parallel Corpora. Language Resources and Evaluation – LREC. 3481-3484.

¹⁴³Seljan, S., Agić, Ž., Tadić, M. (2008). Evaluating sentence alignment on Croatian-English parallel corpora. FASSBL. 101-108.

teksta, točnije ishoda učenja, prikupljenih sa stranica Filozofskog fakulteta Sveučilišta u Zagrebu.

Prije samog početka korištenja alata *Trados Studio*, tekstualne datoteke su pripremljene za segmentaciju i sravnjivanje. To je učinjeno tako što su provjerene pravopisne i gramatičke pogreške. U istraživanju se koristi mrežni alat *Ispravi.me*¹⁴⁴ za provjeru pravopisnih i gramatičkih pogrešaka. Ovo osigurava da tekstualne datoteke budu pravopisno i gramatički ispravne i prikladne za daljnju obradu zbog boljih rezultata u procesima segmentiranja i sravnjivanja (Seljan i sur., 2007).¹⁴⁵ Osim provjere gramatike i pravopisa, tekst koji je napisani u obliku rečenica pretvoren je u oblik natuknica te se svaka natuknica nalazi u zasebnom redu kako bi se stvorila dosljednost kroz cijele korpuse te se nepotrebni interpunkcijski znakovi uklanjaju.

Trados Studio automatski segmentira izvornu i ciljnu datoteku na temelju automatskih postavki alata i potom vrši proces sravnjivanja segmenata. U slučaju nepravilnog segmentiranja ručno se spajaju ili razdvajaju segmenti. U slučaju datoteka iz istraživanja došlo je do pogrešne segmentacije u slučaju kratice. Kao što je prikazano na slici 1, kod kratice “eng.” gdje nakon točke riječ započinje velikim slovom, ručno se spajaju dva segmenta ishodnog jezika kako bi se ispravilo pogrešno segmentiranje i odradilo pravilno sravnjivanje. U istraživanju je ispravak automatskog segmentiranja samo kod navedene kratice.

Studenti će znati osmisлити i provesti projekt poticanja čitanja.	484	484	Students will be able to develop and implement a reading promotion project.
Studenti će znati razlikovati sustave cenzure i propagande u različitim društvenim i političkim sustavima.	485	485	Students will be able to distinguish between systems of censorship and propaganda in different social and political systems.
Računalna analiza teksta i podataka	486	486	Computational Text and Data Analysis
Primijeniti i analizirati proces, algoritme i aplikacije u analizi velikih količina nestrukturiranih podataka (eng.)	487	487	Apply and analyze the process, algorithms and applications in the analysis of large amounts of unstructured data - BIG DATA in order to create new knowledge
Big Data) u cilju stvaranja novoga znanja	488	488	Classify types of structured data and apply them in data processing
Klasificirati vrste strukturiranih podataka i primijeniti ih u obradi podataka	489	489	Apply tools, techniques and methods for statistical analysis of structured data in through the selected software package
Primijeniti alate, tehnike i metode za statističku analizu strukturiranih podataka u kroz odabrani programski paket	490	490	Konstruirati algoritme u skriptnom jeziku (Python, R)
Konstruirati algoritme u skriptnom jeziku (Python, R)	491	491	Compare data visualization tools and methods

Slika 1: Ručno segmentiranje

Kod automatskog sravnjivanja datoteka *Trados Studio* samostalno spaja segmente izvornog i ciljnog jezika kako bi se stvorila prijevodna memorija. Na slici 2 su prikazane veze između segmenata na hrvatskom jeziku i segmenata na engleskom jeziku. Ove veze mogu biti

¹⁴⁴ <https://ispravi.me/>

¹⁴⁵ Seljan, S., Gašpar, A. & Pavuna, D. (2007). Sentence Alignment as the Basis for Translation Memory Database. INFUTURE2007: “Digital Information and Heritage”. 299-311. str. 301, 302

označene različitim bojama kako bi se prikazala razina podudaranja između prijevoda. Veze su označene zelenom, žutom ili crvenom bojom, gdje zelena predstavlja najbolje podudaranje, a crvena najlošije. Podudaranje ovisi o postavkama alata koji se koristi te o jezičnom paru. Ovaj sustav bojanja veza između segmenata pomaže vizualno prikazati razinu podudaranja između prijevoda na hrvatskom i engleskom jeziku. Na temelju tih boja, korisnici mogu brzo prepoznati segmente koji zahtijevaju pažnju ili poboljšanje, te prilagoditi prijevod kako bi postigli veću kvalitetu i točnost. Cilj automatskog sravnjivanja je imati što više zelenih, a što manje crvenih veza. U ovom istraživanju se provjeravalo jesu li segmenti dobro sravnjeni te se nije mijenjao prijevod segmenata.

ishod_hrv_ishod_eng_sdsain		ishod_hrv_ishod_eng_sdsain
Informacijske znanosti	1	1 Information sciences
identificirati i usporediti vodeće trendove i razvoj računalnih tehnologija te tehnologije naprednog pretraživanja	2	2 interpret and solve informational needs of users
demonstrirati vještine stručne usmene i pismene komunikacije	3	3 find, evaluate, and apply information and/or sources of information
interpretirati probleme vezane za dizajn, razvoj, evaluaciju i realizaciju dijaloških sustava utemeljenih na prirodnom jeziku	4	4 apply insights from theoretical and methodological knowledge of information and communication sciences in the solving of professional problems
interpretirati i riješiti informacijske potrebe korisnika	5	5 implement and evaluate archival material and data protection programs
odabrati i upotrijebiti informatičke, informacijske i logičke modele u informacijskim i komunikacijskim znanostima	6	6 select and use methods and techniques of document processing and organization in physical and digital collections and databases
demonstrirati poznavanje osnovnih aktivnosti upravljanja zbirakama predmeta baštine i informacija o njima	7	7 participate in projects in the field of information and communication sciences
interpretirati i koristiti stručnu terminologiju na hrvatskom i engleskom jeziku	8	8 apply, critically evaluate, and connect knowledge in complex information systems management
primijeniti i izgraditi baze podataka i sustave za pretraživanje informacija u informacijskom sustavu	9	9 apply and build databases and information search systems in an information system
primijeniti postupke i načela obrade multimedijских elemenata	10	10 recognize, describe, and connect information offices, systems, and services
pronaći, vrednovati i primijeniti informacije i/ili informacijske izvore	11	11 apply knowledge from different fields of computer processing of language
sudjelovati na projektima iz područja informacijskih i komunikacijskih znanosti	12	12 recognize and apply research results in the field of information and communication sciences
odabrati i upotrijebiti metode i tehnike obrade i organizacije dokumenata u fizičkim i digitalnim zbirakama i fondovima	13	13 select and use information technology, informational, and logical models in information and communication sciences
prikazati funkcioniranje informacijskih institucija	14	14 interpret problems related to design, development, evaluation and realization of dialogic systems based on a natural language

Slika 2: Sravnjivanje datoteka

Nakon sravnjivanja datoteka, spojeni parovi prijevoda se spremaju u prijevodnu memoriju. Iako korpusi sadrže 600 linija teksta, prijevodna memorija sadrži ukupno 535 prijevodnih jedinica. Razlog tome je što u prijevodnoj memoriji ne postoje dva ista segmenta, točnije ako postoje isti segmenti s istim prijevodima oni se spremaju u prijevodnu memoriju samo jednom.

4.1.2. Stvaranje terminološke baze

Terminološka baza u alatu *Trados Studio* stvara se koristeći alat *Multiterm* koji stvara automatske terminološke baze iz prijevodnih memorija, ali samo u slučajevima kada se prijevodne memorije sastoje od minimalno 10 000 prijevodnih jedinica. Međutim, postoji drugi način stvaranja terminoloških baza, a to je ručno izvlačenje termina. Stvorena je terminološka baza ručnim izvlačenjem relevantnih termina iz prijevodne memorije. Termini su inicijalno organizirani u *Microsoft Excel* tablici i izvučeno je 70 terminoloških parova. Osim ručnog izvlačenja termina, moguće je provesti i automatsku ekstrakciju terminologije primjenom odabranih alata temeljenih na jezično neovisnim postavkama (Seljan i sur., 2009¹⁴⁶, Seljan i sur. 2017¹⁴⁷) ili primjenom lokalnih jezičnih gramatika (Seljan i Gašpar, 2009¹⁴⁸; Seljan i sur., 2013¹⁴⁹), no uz obaveznu kasniju reviziju i evaluaciju.

Nakon stvaranje tablice potrebno je pretvoriti terminološku bazu u oblik pristupačan za *Trados Studio* pri čemu se *Microsoft Excel* datoteka pretvara u oblik *MultiTerm* baze. Nakon stvaranja terminološke baze u obliku podržanom u alatu *Trados Studio*, terminološka baza se može koristiti u budućim koracima.

4.1.3. Integracija strojnog prevođenja

U istraživanju se *Trados Studio* povezao s alatom *Google Prevoditelj* (engl. *Google Translate*) kroz API protokol. Kako bi se to omogućilo, potrebno je preuzeti programski dodatak (engl. *plug-in*) koji omogućuje povezivanje s alatom *Google Prevoditelj*. Nakon preuzimanja, novi resurs se dodaje u postavkama projekta. *Google Prevoditelj* se koristi tako da uključuje prijedloge strojnog prevođenja s drugim prijedlozima koje korisnik dobiva za

¹⁴⁶Seljan, S., Dalbello Bašić, B., Šnajder, J., Delač, D., Šamec-Gjurin, M. & Crnec, D. (2009). Comparative Analysis of Automatic Term and Collocation Extraction. *INFuture 2009: Digital resources and knowledge sharing*. 219-228.

¹⁴⁷Seljan, S., Stančić, H. & Dunder, I. (2017). Extracting Terminology by Language Independent Methods *Forum Translationswissenschaft: Translation Studies and Translation Practice* (19). 141-147.

¹⁴⁸ Seljan, S. & Gašpar, A. (2009). First steps in term and collocation extraction from english-croatian corpus. 8th Int. Conf. on Terminology and Artificial Intelligence.

¹⁴⁹Seljan, S., Dunder, I. & Gašpar, A. (2013). From digitisation process to terminological digital resources. *MIPRO 2013*, 1053-1058.

svaki segment. Prijedlog strojnog prevođenja se prikazuje kada nema potpunog ili nejasnog podudaranja te se prijedlog može koristiti kao polazna točka za prijevod (Zaretskaya i sur., 2015¹⁵⁰; Bertoldi i sur., 2013¹⁵¹).

4.2. Testiranje sustava

Za potrebe testiranja sustava, stvorena je tekstualna datoteka sličnog sadržaja kao i stvorena prijevodna memorija. Nova tekstualna datoteka također je iz područja obrazovanja i odnosi se na ishode učenja informacijskih znanosti, te je preuzeta sa stranica Odjela za informacijske znanosti, Sveučilišta u Zadru. Tekstualna datoteka koja se koristi u istraživanju sastoji se od 60 segmenata. Svaki segment predstavlja pojedinačnu izjavu koja je oblikovana u obliku natuknica. Također, cijeli tekst datoteke je pravopisno i gramatički provjeren. Tijekom istraživanja navedena tekstualna datoteka se prevodi na engleski jezik koristeći prijevodnu memoriju, terminološku bazu i automatsko strojno prevođenje.

4.2.1. Analiza izvješća alata Trados Studio

Prema izvješću alata *Trados Studio* prikazanog na slici 3, od ukupno 60 postojećih segmenata njih 6 ima 100% podudaranje s postojećom prijevodnom memorijom. Stoga 3,03% novog teksta je već prethodno prevedeno. Nadalje, postoje 2 segmenta s kontekstualnim podudaranjem u prijevodnoj memoriji, što iznosi 1,60%. Prema izvješću postoje ukupno 52 nova segmenta koji zahtijevaju novi prijevod, točnije 95,37% teksta. Ova informacija sugerira da će veći broj segmenata biti potrebno prevesti zbog nedostatka podudaranja s postojećom prijevodnom memorijom.

¹⁵⁰ Zaretskaya, A., Corpas Pastor, G. & Seghiri M. (2015). Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. *Journal of Translation and Interpretation*. 8(1). 76-88. str. 77

¹⁵¹ Bertoldi, N., Cettolo, M. & Federico, M. (2013). Cache-based Online Adaptation for Machine Translation Enhanced Computer Assisted Translation. *Proceedings of the XIV Machine Translation Summit*. 35-42.

Type	Segments	Words	Characters	Percent	Placeables	Repaired words	Fragment words (whole TU)	Fragment words (TU fragment)	Tags
PerfectMatch	0	0	0	0.00%	0	0	0	0	0
Context Match	2	9	82	1.60%	0	0	0	0	0
Repetitions	0	0	0	0.00%	0	0	0	0	0
Cross-file Repetitions	0	0	0	0.00%	0	0	0	0	0
100%	6	17	137	3.03%	0	0	0	0	0
95% - 99%	0	0	0	0.00%	0	0	0	0	0
85% - 94%	0	0	0	0.00%	0	0	0	0	0
75% - 84%	0	0	0	0.00%	0	0	0	0	0
50% - 74%	0	0	0	0.00%	0	0	0	0	0
New/AT	52	535	4157	95.37%	0	0	8	0	0
Total	60	561	4376	100%	0	0	8	0	0

Slika 3: Izvješće alata *Trados Studio*

Analizom je utvrđeno da od ukupno 60 segmenata, čak 47 segmenata sadrže termine iz stvorene terminološke baze. Broj termina koji se nalaze u izvornom tekstu, a imaju podudaranje s terminološkom bazom je 78, s naglaskom da se neki termini ponavljaju više puta. Korištenje istih termina osigurava točnost i dosljednost u prevođenju tekstualne datoteke (Seljan, 2018)¹⁵² što se može provjeravati naknadno u CAT alatima (Seljan i sur., 2021)¹⁵³ ili u gotovim tekstovima (Gašpar i sur., 2022)¹⁵⁴.

4.2.2. Proces prevođenja

Proces prevođenja kreće od prvog prema posljednjem segmentu. Prijevodi preuzeti iz prijevodne memorije su automatski potvrđeni. Za one segmente za koje ne postoji prijevod u prijevodnoj memoriji, alat *Trados Studio* je generirao prijedlog korištenjem strojnog prevođenja. Korisnik ima mogućnost prihvatiti prijedlog, preurediti ga ili ga u potpunosti odbaciti i stvoriti novi prijevod. Nadalje, korisnik može pretraživati prijevodnu memoriju za određene termine ili dijelove rečenica preko opcije pretrage konkordancije (engl. *concordance*

¹⁵² Seljan, S. (2018), Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. European Parliament: Translation Services in the Digital World, 92-145.

¹⁵³ Seljan, S., Škof Erdelija, N., Kučič, V., Dundđer, I. & Pejić Bach, M. (2021). Quality Assurance in Computer-Assisted Translation in Business Environments. Natural Language Processing for Global and Local Business. 242-270.

¹⁵⁴ Gašpar, A., Seljan, S. & Kučič, V. (2022.) Measuring Terminology Consistency in Translated Corpora: Implementation of the Herfindahl-Hirshman Index. *Information* 13 (2). 43. str. 2

search) koja pretražuje konkordancije odnosno ponavljanja riječi ili fraza u izvornim i ciljanim tekstovima za bolje razumijevanje teksta. Alat za pretraživanje konkordancije ugrađen je u alat *Trados Studio*, iako postoje i online alati koji se mogu koristiti nezavisno (Jaworski i sur., 2021¹⁵⁵)

U istraživanju, ispravci strojnog prevođenja bili su prisutni u 4 prijevoda od 52 prijedloga strojnog prevođenja. Prema tome bilo je potrebno ispraviti 7% prijedloga prijevoda strojnog prevođenja, što je rezultiralo značajnom uštedom vremena u procesu prevođenja. Pogreške u prijedlogu strojnog prevođenja su bile leksičke te se ispravljaju kako bi se bolje prenijelo značenje rečenica. Prikaz ispravaka strojnog prevođenja je prikazan u tablici 1.

Tablica 1: Ispravci prijedloga strojnog prevođenja

	prijedlog strojnog prevođenja	ispravak prijedloga strojnog prevođenja
1.	to understand the interdisciplinary <i>character</i> of information and communication sciences and the strength of connections with related disciplines	to understand the interdisciplinary <i>nature</i> of information and communication sciences and the strength of connections with related disciplines
2.	apply basic terminology <i>in solving</i> related to different information professions and experts	apply basic terminology <i>in addressing issues</i> related to different information professions and experts
3.	Information <i>search</i>	Information <i>retrieval</i>
4.	master the basic concepts of information <i>search</i>	master the basic concepts of information <i>retrieval</i>

Međutim, važno je da korisnici budu pažljivi prilikom korištenja strojnog prevođenja kako ne bi propustili nepravilno prevedene segmente (Zaretskaya i sur., 2015¹⁵⁶; Biçici i sur., 2008¹⁵⁷). Nakon što korisnik potvrdi prijevod segmenta, on se automatski sprema u prijevodnu memoriju koja je korištena tijekom prevođenja. Stoga prijevodna memorija, nakon prevođenja, sadrži 594 prijevodnih jedinica.

¹⁵⁵ Jaworski, R., Duđer, I., Seljan, S. (2021). Usability Analysis of the Concordia Tool Applying Novel Concordance Searching. *Information Technology and Systems*. 1, 128-138.

¹⁵⁶ Zaretskaya, A., Corpas Pastor, G. & Seghiri M. (2015). Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. *Journal of Translation and Interpretation*. 8(1). 76-88. str. 76

¹⁵⁷ Biçici, E. & Dymetman, M. (2008). Dynamic translation memory: Using statistical machine translation to improve translation memory fuzzy matches. In *Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science*. 4919. 454–465. str. 464, 465

4.2.3. Osiguravanje kvalitete prijevoda

Osiguravanje kvalitete prijevoda započelo je stvaranjem tekstualnih datoteka prilikom provjere pravopisa i gramatike te pretvaranjem u oblik natuknice. Postoje različiti alati za provjeru kvalitete prijevoda, koji mogu biti integrirani u CAT alate ili samostalni, komercijalni ili besplatni te nude različite opcije za provjeru kvalitete (Seljan i sur., 2022¹⁵⁸). Nadalje, procjena kvalitete prijevoda integrirana je u alat *Trados Studio*. Kada se otvori opcija procjene kvalitete prijevoda, automatski se prikazuju pogreške. U ovom istraživanju, nakon prevođenja tekstualne datoteke nije prikazana nikakva pogreška. Osim integrirane opcije za procjenu kvalitete prijevoda u alatu *Trados Studio*, postoji nekoliko vanjskih alata koji mogu procijeniti kvalitetu prijevoda prikazom pogrešaka.

U ovom istraživanju je korišten alat *Verifika*. Alat je preuzet kao samostalni program na računalo. Kroz testiranje sustava, prijevod stvoren u alatu *Trados Studio* se uvozi u alat *Verifika* zbog prikaza mogućih pogrešaka tijekom prevođenja.

4.2.3.1. Kategorije pogrešaka

Nakon pregleda prijevoda, postavljaju se postavke provjere pogrešaka u alatu *Verifika*. Postavke unutar alata *Verifika* podijeljene su na:

1. Zaštita (engl. *protection*)
2. Opće postavke (engl. *common*)
3. Dosljednost (engl. *consistency*)
4. Terminologija (engl. *terminology*)
5. Pravopis i gramatika (engl. *spelling and grammar*)
6. Korisnički definirane provjere (engl. *user-defined checks*)
7. Korisnički definirano ignoriranje (engl. *user-defined ignores*)

¹⁵⁸ Seljan, S., Škof Erdelija, N., Kučiš, V., Dunder, I. & Pejić Bach, M. (2021). Quality Assurance in Computer-Assisted Translation in Business Environments. *Natural Language Processing for Global and Local Business*. 242-270.

Postavke zaštite definirane su za svaki tip datoteka, u slučaju ovog istraživanja to je SDL XLIFF datoteka. Ovdje se definiraju vrste prijevodnih jedinica, prevedenih preko alata *Trados Studio*, za koje *Verifika* ne radi provjeru pogrešaka. U vrstu prijevodnih jedinica mogu biti uključene prijevodne jedinice sa 100% podudaranjem, kontekstualnim podudaranjem, prihvaćeni prijevodi, ne prevedeni segmenti i slično. Tijekom istraživanja nije bilo zaštite nad bilo kojom vrstom prijevodnih jedinica.

Opće postavke se dijele na propuste, neprevodive dijelove, zabranjene riječi, velika i mala slova, interpunkcijske znakove i razmake, navodnike i apostrofe te razno. Propusti u prijevodnoj jedinici se odnose na: neprevedene segmente, istog teksta u izvornom i ciljnom segmentu, djelomično prevedenih segmenata i nejednakosti u broju rečenica izvornog i ciljnog segmenta. Neprevodivi dijelovi se većinom koriste kada se u stručnom tekstu termini ili dijelovi rečenice ne prevode na druge jezike kao na primjer riječ *Windows*, te se oni mogu uvesti u alat *Verifika* kako se ne bi prikazivala pogreška. U zabranjene riječi korisnik može dodati riječi koje želi zabraniti u tekstu. Stoga alat *Verifika* provjerava postoje li te riječi u tekstu i prikazuje ih kao pogrešku. Postavke velikih i malih slova provjeravaju pravilno korištenje velikih i malih slova. Stoga *Verifika* prikazuje pogrešku ako je malo slovo umjesto velikog na početku rečenice, kada se veliko slovo nalazi nakon malog slova te se mogu dodati riječi za koje korisnik ne želi da *Verifika* prijavljuje pogrešku, primjer toga može biti često korištena kratica. Interpunkcijski znakovi i razmaci javljaju pogrešku kada se pojavljuje više interpunkcijskih znakova na istom mjestu, kada su nepotrebni razmaci oko interpunkcijskih znakova i kada postoji više od jednog razmaka na istom mjestu. Navodnici i apostrofi provjeravaju postojanje individualnih navodnika i apostrofa bez njihovog para. Pod razno spadaju pogreške neočekivanih znakova koji se ne koriste u jeziku ciljnog teksta, nepotrebno ponavljanje riječi ili fraza jedna za drugom u prijevodu i nejednakosti u dužini izvornog i ciljnog segmenta.

Postavke dosljednosti odnose se posebno na izvorni i posebno na ciljni jezik. Za izvorni jezik provjerava se postoji li više istih segmenata izvornog jezika koji imaju različite prijevode na ciljnom jeziku, a za ciljni jezik se provjerava ima li više istih prijevoda s različitim izvornim segmentima. Mogu se odabrati elementi koji se žele ignorirati tijekom provjere dosljednosti. To uključuje razmake i interpunkciju, velika i mala slova te oblik riječi množine (samo za engleski jezik). Stoga, bez obzira na razlike u razmacima, interpunkciji, velikim i malim slovima, prikazuje se pogreška ako su segmenti isti.

Postavke terminologije provjeravaju postojanje termina u izvornom tekstu za koji ne postoji prijevod u ciljnom tekstu. Ako u izvornom tekstu postoji određen broj termina isti broj termina se treba nalaziti u ciljnom tekstu i obrnuto.

Unutar postavki pravopisa i gramatike identificiraju se moguće pogreške u pravopisu i gramatici prijevoda. Za pogreške pravopisa se prikazuju prijedlozi zamjena za krivo pravopisno napisane riječi. Nadalje, prema postavkama ignoriraju se riječi koje se nalaze u istom obliku u izvornom tekstu, riječi koje sadrže broj u sebi i ne provjeravaju se riječi koje se sastoje od 1 znaka. Za gramatiku se ne prikazuje prijedlog ispravka gramatičke pogreške ili razlog gramatičke pogreške. Pravopis i gramatika se provjeravaju na temelju automatskih postavki jezičnog para koji je korišten, a u istraživanju korišten je hrvatski (HR) i engleski (SAD) jezik.

Korisnički definirane provjere omogućuju korisnicima da definiraju prilagođene provjere s njihovim specifičnim zahtjevima. Korisnici mogu postaviti provjere za određene fraze ili izraze koji su relevantni za njihove prevoditeljske projekte. Korisnički definirano ignoriranje omogućuje korisnicima da navedu elemente ili fraze koje treba zanemariti tijekom provjere kvalitete. To se koristi kod određenog sadržaja za koji se ne radi uobičajenih provjera. U ovom istraživanju nije bilo korisnički definirane provjere i korisnički definiranog ignoriranja.

4.2.3.2. Prikaz alatom pronađenih pogrešaka

Izvješće provjere u alatu *Verifika* prikazuje sve pogreške pronađene u dokumentu prema kategoriji pogreške. Izvješće kod testiranja sustava je pronašlo 3 pogreške. Prve dvije potječu iz općih postavki, dok treća pogreška dolazi iz postavki pravopisa. Prva pogreška, prikazana na slici 4, ukazuje na pogrešku nejednakosti broja rečenica u izvornom i ciljnom tekstu. *Verifika* prikazuje pogrešku u segmentu koji sadrži kraticu *i.e.* Prema interpunkcijskim znakovima alat *Verifika* segment na slici 4 dijeli na više rečenica, stoga ovo nije pogreška nejednakosti broja rečenica. Ova situacija naglašava važnost ručnog pregleda pogrešaka i intervencije kod neispravnih prepoznavanja pogrešaka. U ovom primjeru, pogreška je odbačena. Druga pogreška, isto iz općih postavki, prikazuje pogrešku u razmacima. Na kraju segmenta se nalazi nepotreban razmak.

Source segment	Croatian
prikazati znanja i vještine u pronalaženju informacija s ciljem rješenja konkretnih problema ili daljnjeg učenja, odnosno usavršavanja	
Target segment	English (United States)
demonstrate knowledge and skills in finding information with the aim of solving specific problems or further learning, i.e. improvement	

Slika 4: Prikaz pogreške nejednakosti broja rečenica

Treća pogreška se odnosi na pravopisnu pogrešku. Na slici 5 alat *Verifika* je prikazao riječ "multithreaded" kao pravopisno pogrešno napisanu. *Verifika* daje prijedlog zamjene, a to je "multi threaded". Međutim, unutar konteksta informacijskih znanosti, pravilan prijevod je "multithreaded", stoga je ova pogreška odbačena.

Source segment	Croatian
implementirati osnovna načela višenitnog programiranja u aplikacijama koje zahtijevaju istovremeno izvršavanje na dijeljenim resursima	
Target segment	English (United States)
implement the basic principles of multithreaded programming in applications that require concurrent execution on shared resources	

Slika 5: Prikaz pravopisne pogreške

Nakon analize izvješća iz alata *Verifika*, možemo zaključiti da postoji samo jedna pogreška, a to je dodatni razmak na kraju segmenta. Ova nesavršenost je ispravljena kako bi se očuvala dosljednost unutar prijevodne memorije. Može se reći da je prijevod izrađen pomoću alata *Trados Studio* donio zadovoljavajuće rezultate, a pri tome su prijevodna memorija, terminološka baza i integracija automatskog strojnog prevođenja pozitivno doprinijele procesu prevođenja.

4.3. Usporedna analiza prijevoda

Pilot istraživanje se provodi nad tradicionalnim prijevodima i prijevodima korištenjem CAT alata tekstualne datoteke od strane 10 ispitanika. Ispitanici su studenti različitih fakulteta i smjerova u Zagrebu s poznavanjem pisanog i govornog engleskog jezika. Tekstualna datoteka se sastoji od 5 rečenica, odnosno 108 riječi koje opisuju preddiplomski smjer Informatičke znanosti na Filozofskom fakultetu Sveučilišta u Zagrebu. Svi ispitanici prevode isti izvorni

tekst prvo na tradicionalan način (engl. *Human translation*, HT), a nakon toga, u razmaku od minimalno 5 dana, uz primjenu alata *Trados Studio*.

Tradicionalno prevođenje je provedeno u *Microsoft Word* dokumentu u kojem je ispitanicima dodijeljen izvorni tekst za prijevod. Ispitanicima nije dopušteno koristiti opcije programa *Microsoft Word* koje prikazuju gramatičke i pravopisne pogreške, niti vanjskih alata koji čine isto. Nadalje, s obzirom na manjak rječnika uskovezanih uz ovu temu, dopušteno je korištenje *Google Prevoditelj* preko preglednika. *Google Prevoditelj* se smio koristiti za prijevod pojedinih riječi s kojima se ispitanik prvi put susreće. Upozorilo se ispitanike u *Google Prevoditelj* ne upisuje više uzastopnih riječi ili cijele rečenice. Nadalje, upozorilo ih se na moguće pogreške u prijevodu automatskog strojnog prevođenja. Opciju korištenja *Google Prevoditelj* je koristilo 3 od 10 ispitanika. Nakon prevođenja, prijevodi na tradicionalan način se spremaju u nove projekte u alat *Trados Studio*. Novi projekti, kojih je 10 za tradicionalno prevođenje, su stvoreni bez uključivanja prijevodnih memorija, terminoloških baza ili integracije strojnog prevođenja. U njima se spaja segment izvornog jezika sa segmentom koji je ispitanik preveo na tradicionalan način. Stoga se dokument proizveden u alatu *Trados Studio* može uvesti u program *Verifika* za prikaz i analizu pogrešaka.

U drugom koraku, svih 10 ispitanika prevodi isti tekst primjenom alata *Trados Studio* uz primjenu izgrađenih resursa (prijevodna memorija, terminološka baza) i integracijom alata za automatsko strojno prevođenje. Za svakog ispitanika je stvoren novi projekt u koji su integrirani ovi alati i resursi. Projekti sadrže kopiju prijevodne memorije koja je stvorena za svakog ispitanika zasebno jer se prijevodi spremaju u nju. U alatu *Trados Studio* tekst je podijeljen na 5 segmenata na temelju 5 rečenica. Kod procesa prevođenja u alatu *Trados Studio* nije bilo podudaranja s prijevodnom memorijom jer se izvorni tekst pilot istraživanja sastoji od rečenica, a prijevodna memorija od natuknica. Doduše, ispitanicima je omogućeno korištenje opcije pretrage konkordancije (engl. *Concordance search*) za pojedine termine koji se nalaze u prijevodnoj memoriji. Ispitanici, njih 7 od 10, su u pretrazi konkordancije pretraživali termine: fakultet, studij, smjer, podatak, bibliotekarstvo i muzeologija. Za termine smjer i bibliotekarstvo nije postojala konkordancija u prijevodnoj memoriji i ovi termini se ne nalaze u terminološkoj bazi. Tijekom prevođenja korisnicima se prikazuju termini koji se nalaze u terminološkoj bazi. Postoji 17 pojavljivanja termina u izvornom tekstu iz terminološke baze, s time da se 4 puta ponavljaju isti termini. Za segmente izvornog teksta dan je prijedlog strojnog prevođenja kojeg su ispitanici pregledavali te po potrebi ispravljali. Nakon prevođenja se

prijevodni spremaju u prijevodnu memoriju, zasebno za svakog ispitanika. Potom se dokument projekta stvoren u alatu *Trados Studio* uvozi u alat *Verifika* u svrhu prikaza i analize pogrešaka.

4.3.1. Kriteriji

Usporedna analiza prijevoda provedena je prema sljedećim kriterijima:

1. vrijeme prevođenja
2. broj i kategorija pogrešaka
3. usporedba sličnosti prijevoda.

U procesu usporedbe između tradicionalnog prevođenja i korištenja CAT alata, mjeri se vrijeme potrebno za svaku od ovih metoda kako bi se dobio dublji uvid u njihovu učinkovitost i praktičnost. U vrijeme prevođenja je uključeno samo vrijeme potrebno za prijevod izvornog teksta u ciljni jezik. U vrijeme prevođenja se ne uključuje dodjeljivanje zadatka, učenje sučelja *Trados Studio* i stvaranje projekta. Smatra se kako na vrijeme može utjecati nepoznavanje stručnih termina te njihovo korištenje unutar domene obrazovanja, razumijevanje izvornog teksta, pronalaženje adekvatnih prijevoda ili ekvivalenta, kao i sam proces prevođenja (Webb, 1998¹⁵⁹). Iako primjena CAT alata u početku zahtijeva više vremena radi izgradnje resursa, nakon nekoliko iteracija vrijeme prevođenja se skraćuje u odnosu na tradicionalni postupak (Brkić i sur., 2008¹⁶⁰).

Nakon prevođenja na tradicionalan način i korištenjem CAT alata projekti iz alata *Trados Studio* se uvoze u alat *Verifika* zbog prikaza pogrešaka. Za svaki prijevod (i tradicionalan i primjenom CAT alata) se prolazi kroz pogreške koje prikazuje alat *Verifika* te se odbacuju krivo prikazane pogreške. Potom se brojačano prikazuju pogreške općih postavki, gramatike i pravopisa, koje su objašnjene u prijašnjem poglavlju, posebno za tradicionalno prevođenje i posebno za prevođenje korištenjem CAT alata. Naposljetku se uspoređuje broj pogrešaka, koji generira alat *Verifika*, u tradicionalnom prevođenju s brojem pogrešaka u prijevodu

¹⁵⁹ Webb, L. E. (1998). Advantages and Disadvantages of Translation Memory: A Cost/Benefit Analysis. Monterey Institute of International Studies. Monterey, CA. str. 13

¹⁶⁰ Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. INFUTURE-2009 Digital Resources and Knowledge Sharing. 353-363. str. 355, 356

korištenjem CAT alata. Potrebno je naglasiti kako se ne gleda točnost prijevoda, već se provjerava prijevod u okviru prisutnih pogrešaka.

Prema kriteriju usporedbe sličnosti prijevoda uspoređuju se prijevodi tradicionalnog prevođenja i prevođenja pomoću CAT alata s referentnim prijevodom za svakog ispitanika. Referentni prijevod je nastao profesionalnim ljudskim prijevodom uz korištenje alata *Trados Studio*. Za usporedbu se koristi online alat *GoTranscript*¹⁶¹. Kod usporedbe tekstova, alat prikazuje broj riječi koji se nalazi u tekstu, broj ispravaka te postotak sličnosti. Nadalje, alat vizualno prikazuje ispravke na samom tekstu te prikazuje koja riječ ili riječi se nalaze u prvom tekstu (referentnom prijevodu) i koji je ispravak njih u drugom tekstu (prijevodu od strane ispitanika). Nadalje, uspoređuje se korištenje termina iz domene obrazovanja. Uspoređuje se korištenje stručnih termina i njihove varijante zasebno za tradicionalno prevođenje i prevođenje uz CAT alate. CAT alati omogućavaju smanjenje razlika između prijevoda različitih autora korištenjem resursa i alata koji su integrirani u njih (Bowker i Fisher, 2010).¹⁶²

4.3.2. Analiza rezultata

4.3.2.1. Vrijeme prevođenja

Unutar tablice 2 prikazuje se vrijeme potrebno za prevođenje na tradicionalan način i prevođenje korištenjem CAT alata *Trados Studio* za svakog ispitanika. Za tradicionalno prevođenje nekim ispitanicima bilo je potrebno 7 minuta, dok je drugima 35 minuta. Kod prevođenja pomoću CAT alata vrijeme varira od 4 do 17 minuta. Prosječno vrijeme svih ispitanika je 17,6 minuta za tradicionalno prevođenje, a za prevođenje korištenjem CAT alata taj broj se uveliko smanjio na 8,6 minuta. U prosjeku, vrijeme prevođenja uz pomoć CAT alata u usporedbi s tradicionalnim prevođenjem se smanjilo za 49%. Potrebno je napomenuti kako se u vrijeme prevođenja uključuje samo vrijeme potrebno za stvaranje prijevoda.

U grafikonu 1 vizualno je prikazano vrijeme prevođenja za tradicionalno prevođenje i prevođenje pomoću CAT alata za svakog ispitanika. Na desnoj strani grafikona su prikazana vremena potrebna za prevođenje, izražena u minutama, dok se ispod grafikona nalaze brojevi

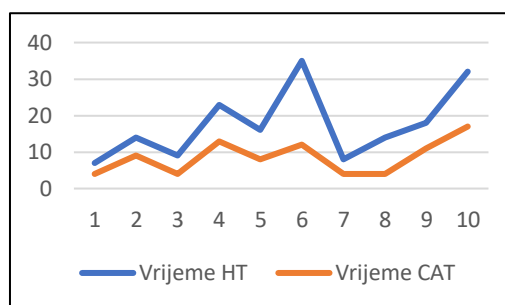
¹⁶¹ <https://gotranscript.com/text-compare>

¹⁶² Bowker, L. & Fisher, D. (2010). Computer-aided translation. Y. Gambier, & L. van Doorslaer (Ur.), Handbook of Translation Studies. 1. 60-65. str. 63

koji označavaju ispitanike. Na grafikonu je jasno vidljivo kako je vrijeme kod tradicionalnog prevođenja uvijek duže od vremena potrebnog za prevođenje pomoću CAT alata.

Tablica 2: Usporedba vremena prevođenja

	vrijeme HT u minutama	vrijeme CAT u minutama	vrijeme HT-vrijeme CAT u postocima
1.	7	4	57%
2.	14	9	64%
3.	9	4	44%
4.	23	13	57%
5.	16	8	50%
6.	35	12	34%
7.	8	4	50%
8.	14	4	29%
9.	18	11	61%
10.	32	17	53%
prosjeak svih ispitanika	17,6	8,6	49%



Grafikon 1: Usporedba vremena prevođenja

Postoji znatna razlika u vremenu prevođenja između tradicionalnog prevođenja i prevođenja pomoću alata *Trados Studio*. Na ovu razliku može utjecati nepoznavanje ispitanika sa stručnom terminologijom i njezino korištenje unutar rečenice. Primjer nepoznavanja je zbunjenost ispitanika kod korištenja termina poput *studij*, *bibliotekarstvo*, *muzeologija*, itd. Nadalje, na kraće vrijeme prevođenja kod korištenja CAT alata može utjecati prikaz strojnog prevođenja. Stoga ispitanici ne moraju prevoditi cijele rečenice, već samo pregledavati prijedloge te ih po potrebi ispravljaju. Naposljetku, na vrijeme prevođenja u CAT alatu može utjecati i prijašnje tradicionalno prevođenje jer su ispitanici bili već upoznati s tekstom koji se prevodi, a kod tradicionalnog prevođenja ispitanici se prvi put susreću s njime.

4.3.2.2. Broj i kategorija pogrešaka

Tablica 3 prikazuje broj pogrešaka nastalih kroz tradicionalno prevođenje koji je generirao alat *Verifika*. U tablici 3, za svakog ispitanika, prikazane su pogreške za opće

postavke, pravopis i gramatiku u prijevodu tradicionalnog prevođenja te njihov sveukupni broj. Od sveukupnih 10 ispitanika, njih pola (5) nema niti jednu pogrešku. Najveći broj pogrešaka po ispitaniku je 21. Općih pogreška je sveukupno dvije, pravopisnih pogrešaka je 31, a gramatičkih 6. Sveukupni broj pogrešaka od svih ispitanika je 39 te se može primijetiti veliki broj pogrešaka s obzirom na mali opseg teksta koji se prevodio.

Tablica 3: Broj pogrešaka tradicionalnog prevođenja

	opće postavke	pravopis	gramatika	sveukupan broj pogrešaka
1.	0	0	0	0
2.	0	0	1	1
3.	0	0	0	0
4.	1	18	2	21
5.	0	0	0	0
6.	0	3	0	3
7.	0	0	0	0
8.	0	0	0	0
9.	0	4	1	5
10.	1	6	2	9
sveukupan broj pogrešaka	2	31	6	39

Prva pogreška općih postavki se odnosi na broj razmaka, gdje je u prijevodu izvornog segmenta dvostruki razmak unutar rečenice. Druga pogreška se odnosi na nedosljednost brojeva u izvornom i ciljnom segmentu. U izvornom segmentu se ne nalazi brojka jedan, a u prijevodu se nalazi *one*:

- Izvorni segment: Ovaj smjer obuhvaća raznolike teorijske i istraživačke pristupe relevantne za osnove informacijskih i komunikacijskih znanosti.
- Prijevod: This *one* direction includes varied theoretical and research join relevant for the basics informational and communicational science.

Osim navedenih pogrešaka općih postavki pojavilo se 10 pogrešaka, kod svakog ispitanika jedna, koje se odnose na nedosljednost brojeva u izvornom i ciljnom jeziku. Ove pogreške su odbačene:

- Izvorni segment: Završetkom preddiplomskog studija studenti su osposobljeni za daljnje obrazovanje putem diplomskih studija kao što su arhivistika, bibliotekarstvo, muzeologija te *dvije* grane informatike – istraživačka i nastavnička.

- Prijevod: Upon completion of undergraduate studies, students are qualified for further education through graduate studies such as archival science, library science, museology and *two* branches of informatics - research-oriented and teaching-oriented.

Pravopisnih pogrešaka je najviše, čak 31, gdje su pogreške pravopisno krivo napisanih riječi. U tablici 4 su prikazane pravopisne pogreške, broj ponavljanja iste pogreške i njihovi ispravci. Ove pogreške se pojavljuju zbog nepoznavanja točnog pravopisa riječi ili nepoznavanja samog termina. Nadalje, u riječima *information* i *data* korisnici nisu koristili pravilnu množinu riječi u engleskom jeziku te su stavljali slovo *s* na kraj riječi.

Tablica 4: Pravopisne pogreška tradicionalnog prevođenja

	pogreška	broj ponavljanja	ispravak
1.	analising	1	analysing / analyzing
2.	archivistics	1	archival science
3.	arhiviests	1	archival science
4.	approachs	1	approaches
5.	comperhensive	1	comprehensive
6.	datas	3	dana
7.	graduted	1	graduated
8.	includong	1	including
9.	informations	5	information
11.	informetic	1	information
12.	methodes	1	methods
13.	organisating	2	organizing
14.	preceptorial	1	preceptorial
15.	programm	1	program / programme
16.	reasearch	2	research
17.	recodnise	1	recognise
18.	surten	1	certain
19.	tehnologies	1	technologies
20.	tehnology	3	technology

21.	undergraduated	1	undergraduate
22.	usege	1	usage

Osim navedenih pogrešaka, alat *Verifika* je riječi *analysing* i *organisation* prikazao kao pravopisno pogrešno napisane riječi. Razlog ovih pogrešaka je korištenje engleskog jezika korištenog u Sjedinjenim Američkim Državama, te se prema tome pišu *analyzing* i *organization*. Za riječ *analysing* alat *Verifika* je pokazala ponavljanje 2 puta, a za riječ *organisation* 5 puta. Doduše ove pogreške su odbačene jer su riječi pravopisno točno napisane u oba ekvivalenta.

Gramatičke pogreške, kojih je 6, su pogreške ovisne o gramatičkim pravilima engleskog jezika. U tablici 5 su prikazane navedene pogreške koje je generirao alat *Verifika*. U tablici je prikazan i ispravak navedenih gramatičkih pogrešaka. Dio rečenice koji je označen je dio koji je alat *Verifika* označio kao gramatičku pogrešku, ali ispravci se događaju i u riječima oko tog dijela.

Tablica 5: Gramatičke pogreške tradicionalnog prevođenja

	pogreška	ispravak
1.	Also, <i>student are</i> able to...	Also, <i>students are</i> able to...
2.	<i>After graduating</i> undergrad study...	<i>After graduating</i> from undergrad study.
3.	Onwards, <i>main focus</i> on...	Onwards, <i>the main focus</i> on...
4.	...study <i>students are</i> empowerd...	...study, <i>students are empowered</i> ...
5.	This one <i>direction includes</i> diverse...	This <i>direction includes</i> diverse...
6.	... <i>offer comprehensive</i> understanding...	... <i>offer a comprehensive</i> understanding...

Alat *Verifika* je prikazao 9 gramatičkih pogrešaka koje su odbačene, a prikazane su u tablici 6. Alat *Verifika* ne navodi razlog gramatičke pogreške i ne daje prijedloge ispravka. Detaljnom analizom označenog dijela i rečenice u kojoj se nalazi nije uočena gramatička pogreška.

Tablica 6: Odbačene gramatičke pogreške tradicionalnog prevođenja

	pogreška	broj ponavljanja	mogući ispravci
1.	This <i>course includes</i> diverse...	2	This <i>course covers</i> diverse...
2.	...students <i>will learn</i> how to...	1	/
3.	Furthermore, <i>focus is</i> given	2	/
4.	...at <i>the Faculty</i> of...	2	/
5.	However, <i>students are</i> going to...	1	/
6.	Also, <i>it focuses</i> on...	1	/

Za prijevod korištenjem CAT alata, alat *Verifika* je prikazao 10 pogrešaka općih postavki, 0 pogrešaka za pravopis i 10 pogrešaka za gramatiku. Provjerom pogrešaka se ustanovilo da se odbacuju sve pogreške dodijeljene od strane alata *Verifika*.

Pogreške općih postavki, kojih je sveukupno 10, a po ispitaniku 1, su pogreške nedosljednosti brojeva u izvornom i ciljnom segmentu. Ova pogreška se pojavljivala i kod tradicionalnog prevođenja:

- Izvorni segment: Završetkom preddiplomskog studija studenti su osposobljeni za daljnje obrazovanje putem diplomskih studija kao što su arhivistika, bibliotekarstvo, muzeologija te *dvije* grane informatike – istraživačka i nastavnička.
- Prijevod: Upon completion of undergraduate studies, students are qualified for further education through graduate studies such as archival science, library science, museology and *two* branches of informatics - research-oriented and teaching-oriented.

Alat *Verifika* je identificirao 10 gramatičkih pogrešaka u rečenici:

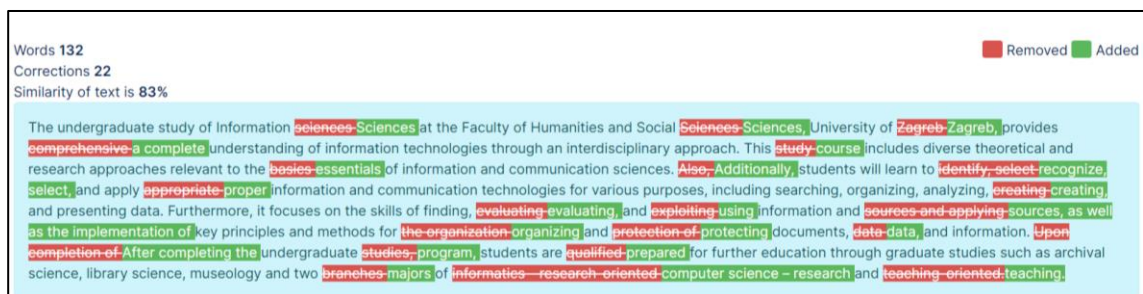
- *This course includes diverse theoretical and research approaches...*

Alat *Verifika* je za svakog ispitanika prikazao jednu pogrešku, što rezultira ukupno 10 gramatičkih pogrešaka na istom mjestu. Ovdje nije uočena gramatička pogreška, ali može se razmotriti ispravak u obliku "*This course covers diverse...*" kako bi rečenica bila razumljivija. Do ponavljanja ove pogreške u prijevodima svih ispitanika događa se zbog korištenja istih resursa i alata te su ispitanici podložni njima. Tijekom prevođenja, ispitanici se oslanjaju na strojno prevođenje i većinom ispravljaju netočne prijevode koje je generirao *Google Prevoditelj*.

Usporedbom broja pogreška u prijevodu tradicionalnog prevođenja i prevođenja pomoću CAT alata uočeno je znatno smanjenje. Tijekom tradicionalnog prevođenja sveukupni broj pogrešaka bio je 39. Prevođenjem pomoću CAT alata je taj broj sveden na 0. Naglašava se kako je provedba osiguranja kvalitete bila provedena kroz alat *Verifika*, te je prikaz i broj pogrešaka ovisan o njegovim rezultatima.

4.3.2.3. Usporedba sličnosti prijevoda

Usporedba sličnosti prijevoda je provedena primjenom alata *GoTranscript*, s ciljem utvrđivanja sličnosti između tradicionalnog prevođenja i prevođenja primjenom CAT alata s referentnim prijevodom. Na slici 6 je prikazan primjer usporedbe sličnosti između referentnog teksta i tradicionalnog prijevoda jednog od ispitanika. *GoTranscript* broji broj zajedničkih riječi, broj ispravaka i postotak sličnosti. Crvenom bojom je označen tekst iz referentnog teksta, a zelenom bojom je prikazano kako je ispitanik preveo.



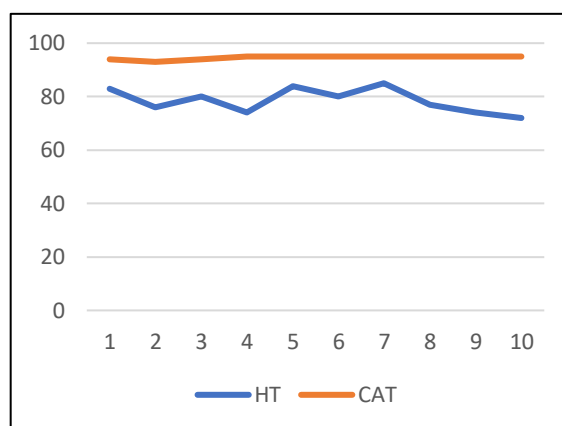
Slika 6: Prikaz usporedbe sličnosti prijevoda u alatu *GoTranscript*

Tablica 7 prikazuje usporedbu sličnosti prijevoda tradicionalnog prevođenja i prevođenja pomoću CAT alata s referentnim prijevodom za svakog ispitanika. Rezultati u tablici 7 pokazuju da postoji manja sličnost u rezultatima tradicionalnog prevođenja u usporedbi s CAT alatom. Konkretno, kod svakog je ispitanika zabilježen niži postotak sličnosti u tradicionalnom prevođenju u odnosu na prevođenje putem CAT alata. Najmanji postotak sličnosti u tradicionalnom prevođenju iznosi 72%, a najveći 85% za razliku od prevođenja pomoću CAT alata gdje je najmanji postotak sličnosti 93% i kreće se do 95%. U prosjeku, postotak sličnosti

za tradicionalno prevođenje iznosio je 79%, dok je za prevođenje pomoću CAT alata taj postotak bio visokih 95%. U grafikonu 2 su vizualno prikazani ovi rezultati usporedbe. Prema grafikonu se prikazuje postotak sličnosti za svakog ispitanika u tradicionalnom prevođenju i prevođenju pomoću CAT alata. Iz grafikona je jasno vidljivo kako je sličnost u prijevodu pomoću CAT alata veća za svakog ispitanika u usporedbi s tradicionalnim prevođenjem. Nadalje, postotak za tradicionalno prevođenje varira, a kod CAT alata je skoro jednaka kod svih ispitanika.

Tablica 7: Usporedba sličnosti prijevoda

	HT	CAT
1.	83%	94%
2.	76%	93%
3.	80%	94%
4.	74%	95%
5.	84%	95%
6.	80%	95%
7.	85%	95%
8.	77%	95%
9.	74%	95%
10.	72%	95%
Prosječno:	79%	95%



Grafikon 2: Usporedba sličnosti prijevoda

Važno je napomenuti da niti jedan od ispitanih sudionika nije postigao stopu sličnosti od 100%, već je najveći postotak bio 95%. Detaljnom analizom prijevoda pomoću CAT alata primijećeno je kako se razlike u prijevodima većinom temelje na krivom prijevodu termina koji se ne nalaze u terminološkoj bazi.

Sveukupni broj termina koji se nalazi u izvornom tekstu, a imaju poklapanje s terminološkom bazom je 17. Od navedenih 17 termina, 4 se termina ponavljaju. Tijekom prevođenja na tradicionalan način korisnici nisu imali pristup terminološkoj bazi i prijevodu u njoj. Osim navedenih termina u izvornom tekstu se nalazi još 11 termina iz domene obrazovanja. U tablici 8 se prikazuju svi termini s pogrešnim prijevodom. U tablici 8 je prikazan termin na hrvatskom jeziku, njegov prijevod na engleski jezik, pogrešni prijevod proizveden od ispitanika i broj ponavljanja. Neki termini su imali različite varijante prijevoda kod prevođenja, a to su *arhivistika*, *bibliotekarstvo*, *informacijska tehnologija*, *istraživačka (informatika)*, *nastavnička (informatika)*, *osposobljen* i *studij* što dovodi do nedosljednosti između prijevoda. Kod termina *istraživačka (informatika)* i *nastavnička (informatika)*

ispitanici su koristili prijevod *research* i *teaching* umjesto *research-oriented* i *teaching-oriented*. Kod ovih prijevoda tekst je i dalje razumljiv te su ove pogreške zanemarive, ali ne i ostale varijante prijevoda.

Tablica 8: Krivi prijevodi termina u tradicionalnom prevođenju

	termin na hrvatskom jeziku	termin na engleskom jeziku	prijevod kod tradicionalnog prevođenja	broj ponavljanja
1.	arhivistika	arhival science	arhivists	1
			arhivism	1
			arhivistics	1
			arhival studies	2
			arhiving studies	1
2.	bibliotekarstvo	library science	librarianship	3
			library studies	1
			libraries studies	1
			librarian studies	1
3.	diplomski (studij)	graduate	postgraduate	1
4.	dokument	document	file	1
5.	informacija	information	informations	1
			dana	1
			informatic	1
6.	informatička tehnologija	information technologies	information technology	1
			informatic technology	1
			technology of informations	1
7.	informatičke i komunikacijske znanosti	informational and communicational sciences	information and communication sciences	3
8.	informatičke znanosti	information sciences	informatic sciences	1
9.	istraživačka (informatika)	research-oriented	research	8
			research course	2

10.	muzeologija	museumology	museums studies	1
			museumology studies	1
			museumology and heritage studies	1
11.	načelo	principle	guides	1
12.	nastavnička (informatika)	teaching-oriented	educational course	1
			education	1
			preceptorial	1
			teaching	5
			teacher	1
			teaching course	1
13.	osposobljen	qualified	empower	1
			capable	3
			prepared	1
14.	podatak	dana	datas	3
15.	preddiplomski (studij)	undergraduate	graduate	2
16.	pristup	approach	method	1
17.	pronalaženje	finding	searching	3
18.	studij	study	program	3
			path	1
			course	5
			major	1
			direction	1
19.	tehnologija	technology	tehnology	1

U tablici 8 se nalaze pogreške koje se pojavljuju zbog pravopisno pogrešno napisanog termina, neispravnosti pisanja množine termina i nepoznavanja stručnog prijevoda termina. Sveukupni broj termina koji imaju pogrešan prijevod je 72. S obzirom na to da ne postoji

određena terminološka baza za tradicionalno prevođenje, kao što je to u CAT alatima, postoje različiti ekvivalenti termina u prijevodu što dovodi do nedosljednosti između tradicionalnog prijevoda između ispitanika i manjoj stopi sličnosti.

U tablici 9 su prikazane pogreške u prijevodu termina kod prevođenja pomoću CAT alata. U tablici 9 je prikazan termin na hrvatskom jeziku, njegov prijevod na engleski jezik, pogrešni prijevod proizveden od ispitanika i broj ponavljanja. Kod prevođenja pomoću CAT alata pogrešno prevedeni prijevodi termina koji se nalaze u terminološkoj bazi su samo za termin *studij* koji se prevodio s *major*, *approach* i *program*. U kontekstu rečenice može se koristiti termin *major*, ali u svrhu dosljednosti i bolje razumijevanje konačnog teksta bolje je koristiti samo jedan termin (Gašpar i sur., 2022¹⁶³). Osim navedenog termina, svi termini iz terminološke baze su pravilno prevedeni. Doduše, tijekom usporedbe sličnosti prijevoda s referentnim prijevodom vidljivi su termini koji su krivo prevedeni kroz više prijevoda. Sveukupni broj pogrešno prevedenih termina je 37. Kod termina *istraživačka (informatika)* i *nastavnička (informatika)* ispitanici su koristili prijevod *research* i *teaching* umjesto *research-oriented* i *teaching-oriented*, ali je tekst i dalje razumljiv te su ove pogreške zanemarive, ali ne i ostale varijante. Osim navedenih termina pogreške se pojavljuju kod termina *bibliotekarstvo* i *pretraživanje*.

Tablica 9: Krivi prijevodi termina u prevođenju pomoću CAT alata

	termin na hrvatskom jeziku	termin na engleskom jeziku	prijevod uz korištenje cat alata	broj ponavljanja
1.	bibliotekarstvo	library sciences	librarianship	7
2.	istraživačka (informatika)	research-oriented	research	10
3.	nastavnička (informatika)	teaching-oriented	teaching	9
			education	1
4.	pretraživanje	searching	retrieving	1
5.	studij	study	major	7
			approach	1
			study program	1

¹⁶³ Gašpar, A., Seljan, S. & Kučiš, V. (2022.) Measuring Terminology Consistency in Translated Corpora: Implementation of the Herfindahl-Hirshman Index. *Information* 13 (2). 43. str. 2

Termin *bibliotekarstvo* je u tradicionalnom prijevodu 6 puta preveden krivo u 4 različite varijante, što dovodi do nedosljednosti prijevoda. Isti termin je 7 puta preveden krivo u prevođenju pomoću CAT alata, ali samo u jednoj varijanti, a to je *librarianship*. Razlog povećanja broja krivo prevedenog termina *bibliotekarstvo* je što automatsko strojno prevođenje pruža prijevod *librarianship* i ispitanici koji nisu upućeni u termin ne mogu znati njegov točan prijevod. Stoga se zaključuje kako integracija automatskog strojnog prevođenja može dovesti do neispravnosti korištenja termina ako oni ne postoje u prijevodnoj memoriji i terminološkoj bazi. Osim termina *bibliotekarstvo* postoji još termina koji se trebaju nadodati u terminološku bazu, a koji nisu imali pogrešku u prevođenju kod CAT alata jer je automatsko strojno prevođenje proizvelo točan prijevod termina. Primjer takvog termina je *arhivistika*. Navedeni termin je bio pogrešno preveden kod 6 ispitanika u tradicionalnom prevođenju u 5 ekvivalenta. Stoga se smatra da je strojno prevođenje u ovom slučaju spriječilo ponavljanje iste pogreške tako što je točno prevelo termin. Međutim, kao što je vidljivo u slučaju termina *bibliotekarstvo*, to nije uvijek slučaj. Stoga se smatra da treba nadopuniti terminološku bazu sa svim relevantnim terminima koji se nalaze u izvornom tekstu, te koji su pogrešno prevedeni kod tradicionalnog prevođenja.

5. Zaključak

Cilj rada bio je ispitati učinkovitost primjene alata za računalno potpomognuto prevođenje (CAT alata) radi utvrđivanja raznolikosti stručne terminologije, vrsta i kategorija pogrešaka i stvaranja prijevoda sukladnog referentnom prijevodu.

U teorijskom dijelu rada analizirana je uloga CAT sustava u procesu prevođenja te su detaljno opisane sastavne komponente cjelokupnog CAT sustava. CAT sustave predstavlja skupina programa koji kombiniraju različite resurse i alate (prijevodna memorija, terminološka baza, integracija strojnog prevođenja) koji pomažu korisnicima kod procesa prevođenja, a zatim prijenosa informacija s jednog prirodnog jezika na drugi što omogućuje brži pristup informacijama.

U istraživačkom dijelu se prikazuju mogućnosti i nedostaci sustava za računalno potpomognuto prevođenje. Izgrađena je prijevodna memorija procesom sravnjivanja korpusa, terminološka baza izvlačenjem relevantnih termina iz prijevodne memorije te se integriralo automatsko strojno prevođenje. U pilot istraživanju, 10 sudionika je prevodilo isti tekst iz domene obrazovanja, specifično informacijskih znanosti, na dva različita načina: tradicionalno i uz pomoć CAT alata. Prvo svi ispitanici prevode tekst na tradicionalan način koristeći *Microsoft Word*, bez pristupa alatima za pravopisnu i gramatičku provjeru, s iznimkom ograničene upotrebe *Google Prevoditelja* za pojedinačne riječi. Nakon toga, isti tekst je preveden pomoću CAT alata *Trados Studio*. Sudionicima je bilo omogućeno pregledavanje prethodnih prijevoda iz prijevodnih memorija, pretraga termina u terminološkoj bazi i analiza prijedloga strojnog prevođenja.

Rezultati istraživanja pokazuju upola kraće vrijeme prevođenja, u prosjeku 49%, korištenjem CAT alata za razliku od tradicionalnog prevođenja, iako su ispitanici isti tekst prevodili dva puta što je zasigurno utjecalo na vrijeme prevođenja. Nadalje, broj pogrešaka se uvelike smanjio korištenjem CAT alata koji daje informaciju korisniku o mogućim pogreškama koje se nalaze u prijevodu. Broj pogrešaka kod tradicionalnog prevođenja se smanjio s 39 na nulu. Najveći broj pogrešaka koji se smanjio je bio u pravopisnim pogreškama, kojih je bilo 31 kod tradicionalnog prevođenja. Razlog tome je što CAT alati imaju ugrađene alate za provjeru pravopisnih i gramatičkih pogrešaka. Osim ovih alata, smatra se kako je automatsko strojno prevođenje također bilo ključno u smanjenju ovih pogrešaka. Usporedbom sličnosti prijevoda se potvrdila veća dosljednost u prevođenju pomoću CAT alata. Kod tradicionalnog prevođenja

sličnost s referentnim tekstom je bila 79%, a korištenjem CAT alata 95%. Funkcionalnosti CAT alata omogućile su ispitanicima snalaženje s terminologijom s kojom nisu bili upoznati. U tradicionalnom prevođenju pogrešno prevedenih termina je sveukupno 72, a kod prevođenja pomoću CAT alata 37. Kako bi se ova brojka još više smanjila, potrebno je nadopuniti terminološku bazu.

Istraživanje ima svoje ograničenosti, uključujući prevođenje istog teksta i tradicionalno i uz pomoć CAT alata, ograničenu količinu prikupljenih resursa, mali broj sudionika te upotrebu samo jednog alata za identifikaciju pogrešaka u prijevodima. Zbog ovih faktora, ovo pilot istraživanje se smatra samo inicijalnim korakom u razumijevanju mogućih prednosti i nedostataka upotrebe CAT alata.

Iako su potvrđene pretpostavke o smanjenju vremena prevođenja i broju pogrešaka te o boljoj dosljednosti korištenjem CAT alata, primijećeni su i nedostaci korištenja. Nedostaci se primjećuju u nepravilnom prijevodu terminologije kod ispitanika. Primjerice termin *bibliotekarstvo* je manje puta pogrešno preveden kod tradicionalnog prevođenja u usporedbi s prevođenjem pomoću CAT alata. Razlog tome je što je automatsko strojno prevođenje taj termin prevelo krivo te ispitanici, koji nisu bili upoznati s terminom i njegovim prijevodom, nisu ispravili tu pogrešku. Smatra se kako su ispitanici koji su i upoznati s terminom previdjeli ovu pogrešku zbog prijevoda danog od strane strojnog prevođenja te se mora naglasiti da CAT alati omogućuju lakše prevođenje, ali je krajnji prijevod ipak na samim korisnicima. Stoga je potrebno nadopuniti terminološku bazu s novim terminima i povećati prijevodnu memoriju kako bi se ove pogreške smanjile.

Brzina i kvaliteta prijevoda koje ovi alati omogućuju pruža značajnu prednost korisnicima. Iako postoji izazov u svladavanju samog alata, ulaganje vremena i truda u učenju istih će dugoročno rezultirati povećanom učinkovitošću i smanjenjem vremena potrebnog za prevođenje.

6. Literatura:

1. Bertoldi, N., Cettolo, M. & Federico, M. (2013). Cache-based Online Adaptation for Machine Translation Enhanced Computer Assisted Translation. Proceedings of the XIV Machine Translation Summit. 35-42.
2. Biçici, E. & Dymetman, M. (2008). Dynamic translation memory: Using statistical machine translation to improve translation memory fuzzy matches. In Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science. 4919. 454–465.
3. Blancafort, H., Heid, U., Gornostay, T., Méchoulam, C., Daille, B. & Sharoff, S. (2011). User-centred Views on Terminology Extraction Tools: Usage Scenarios and Integration into MT and CAT Tools. Tralogy I. Métiers et technologies de la traduction: quelles convergences pour l'avenir?
4. Bowker, L. & Fisher, D. (2010). Computer-aided translation. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 60-65.
5. Braune, F. & Fraser, A. (2010). Improved unsupervised sentence alignment for symmetrical and asymmetrical parallel corpora. *Proc. 23rd International Conf. Computational Linguistics (COLING 2010)*. 81-89. Doi: 10.5555/1944566.1944576
6. Brkić, M., Seljan, S. & Bašić Mikulić, B. (2009). Using Translation Memory to Speed up Translation Process. *INFuture-2009 Digital Resources and Knowledge Sharing*. 353-363.
7. Brkić Bakarić, M., Matetić, M. & Seljan, S. (2011). Towards Obtaining High Quality Sentence-Aligned English-Croatian Parallel Corpus. *Conference: 4th IEEE International Conference on Computer Science and Information Technology ICCSIT 2011.At: Chengdu, China*. 241-248.
8. Çetiner, C. (2021). Evaluating Quality Assurance (QA) Checks in CAT tools in Relation to Professional Quality Standards. *Euroasian Conference on Language and Social Sciences*. 338-349.
9. Chiocchetti, E., Wissik T., Lušický, V. & Wetzel, M. (2017). Quality assurance in multilingual legal terminological databases. *The Journal of Specialised Translation*. 27. 164-188.
10. Choudhury, R., & McConnell, B. (2013). Translation Technology Landscape Report. TAUS – Translation Automation User Society. Nizozemska.

11. Czopik, J. (2014). Quality Assurance process in translation. *Translating and Computer*. 36. 77-85.
12. Dovedan, Z., Seljan, S. & Vučković K. (2002). Strojno prevođenje kao pomoć u procesu komunikacije. *Informatologia*. 35 (4). 283-291.
13. Dunder, I., Seljan, S., Pavlovski, M. (2020). Automatic Machine Translation of Poetry and a Low-Resource Language Pair. *MIPRO*, 1034-1039.
14. Dunder, I., Seljan, S., Pavlovski, M. (2021). What Makes Machine-Translated Poetry Look Bad? A Human Error Classification Analysis. *Central European Conference on Information and Intelligent Systems*. 183-191.
15. Fernández-Parra, M. (2010). The Workflow of Computer-Assisted Translation Tools in Specialised Translation. *Reconceptualizing LSP: Online proceedings of the XVII European LSP Symposium 2009*. URL: <http://bcom.au.dk/fileadmin/www.asb.dk/isek/fernandez-parra.pdf>
16. Folaron, D. A. (2010). Translation tools. Y. Gambier, & L. van Doorslaer (Ur.), *Handbook of Translation Studies*. 1. 429-436.
17. Gale, W. & Church, K. (1993). A program for aligning sentences in bilingual corpora. *Computational linguistics*. 19(1). 75-102. Doi: 10.3115/981344.981367.
18. Gašpar, A., Seljan, S. & Kučiš, V. (2022.) Measuring Terminology Consistency in Translated Corpora: Implementation of the Herfindahl-Hirshman Index. *Information* 13 (2). 43. doi: <https://doi.org/10.3390/info13020043>.
19. Guillou, L. (2013). Analysing Lexical Consistency in Translation. *In Proceedings of the Workshop on Discourse in Machine Translation*. 10–18.
20. Han, B. (2020). Translation, from Pen-and-Paper to Computer-Assisted Tools (CAT Tools) and Machine Translation (MT). *Proceedings 2020*, 63(1), 56. <https://doi.org/10.3390/proceedings2020063056>
21. Hartley, T. (2009). Technology and Translation. Munday, J. (Ur.) *The Routledge Companion to Translation Studies*. 106-127.
22. Hutchins, J. (1986). *Machine Translation: Past, Present, Future*. Chichester: Ellis Horwood.
23. Hutchins, J. (1995). Machine Translator: A Brief History. E. F. K. Koerner & R. E. Asher (Ed.). *Concise History of the Language Sciences*. 431-445.
24. Hutchins, J. (2003). The development and Use of Machine Translation Systems and Computerbased Translation Tools. *International Journal of Translation*. 15(1). 6-26.

25. Irfan, M. (2017). Machine Translation. Department of Computer Science. Bahria University Islamabad. Dostupno 30.8.2023. na:
https://www.researchgate.net/publication/320730405_Machine_Translation
26. ISO 17100:2015. (bez dat.). ISO. <https://www.iso.org/standard/59149.html>
27. Jaworski, R., Duđer, I., Seljan, S. (2021). Usability Analysis of the Concordia Tool Applying Novel Concordance Searching. *Information Technology and Systems*. 1, 128-138.
28. Jaworski, R., Seljan, S., & Dunder, I. (2023). Four Million Segments and Counting: Building an English-Croatian Parallel Corpus through Crowdsourcing Using a Novel Gamification-Based Platform. *Information*. 14 (4). 226. Doi:
<https://doi.org/10.3390/info14040226>
29. Kerremans, K. (2010). Comparative Study of Terminological Variation in Specialised Translation. *Reconceptualizing LSP. XVII Eur. LSP Symp.* 1–14.
30. Kotwicki, M. (2018). Outsmarting yourself. Self-revision and self-impruvment. *TransELTE 2018*. 61-72.
31. Kučič, V., Seljan, S., Klasnić, K. (2009). Evaluation of electronic translation tools through quality parameters. *INFuture 2009: Digital Resources and Knowledge Sharing*. 341-351.
32. Moore, R. (2002). Fast and accurate sentence alignment of bilingual corpora. *Proc. 5th Conf. Association for Machine Translation in the Americas (AMTA 02)*. 135-144.
33. Moran, J., Saam, C. & Lewis, D. (2014). Towards desktop-based CAT tool instrumentation. *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*. 99-112.
34. Petrova, V. (2019). Translation Quality Assessment Tools and Processes in Relation to CAT Tools. *In Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)*. 89-97. Doi: 10.26615/issn.2683-0078.2019_011
35. Seljan, S. (2018). Total Quality Management Practice in Croatian Language Service Provider Companies. *ENTRENOVA '18 - ENTERprise REsearch InNOVAtion Conference*. 4 (1). 461-469.
36. Seljan, S. (2018). Quality Assurance (QA) of Terminology in a Translation Quality Management System (QMS) in the Business Environment. *European Parliament: Translation Services in the Digital World*. 92-145.

37. Seljan, S., Agić, Ž., Tadić, M. (2008). Evaluating sentence alignment on Croatian-English parallel corpora. FASSBL. 101-108.
38. Seljan, S., Dalbelo Bašić, B., Šnajder, J., Delač, D., Šamec-Gjurin, M. & Crnec, D. (2009). Comparative Analysis of Automatic Term and Collocation Extraction. INFUTURE 2009: Digital resources and knowledge sharing. 219-228.
39. Seljan, S., Dunder, I. & Gašpar, A. (2013). From digitisation process to terminological digital resources. MIPRO 2013, 1053-1058. DOI: 10.13140/RG.2.1.1883.9521
40. Seljan, S., Dunder, I., Pavlovski, M. (2020). Human Quality Evaluation of Machine-Translated Poetry. MIPRO, 1040-1045.
41. Seljan, S. & Gašpar, A. (2009). First steps in term and collocation extraction from english-croatian corpus. 8th Int. Conf. on Terminology and Artificial Intelligence.
42. Seljan, S., Gašpar, A. & Pavuna, D. (2007). Sentence Alignment as the Basis for Translation Memory Database. *INFUTURE2007: "Digital Information and Heritage"*. 299-311.
43. Seljan, S., Tadić, M., Agić, Ž., Šnajder, J., Dalbelo Bašić, B. & Osmann, V. (2010). Corpus Aligner (CorAl) Evaluation on English-Croatian Parallel Corpora. Language Resources and Evaluation – LREC. 3481-3484.
44. Seljan, S., Stančić, H. & Dunder, I. (2017). Extracting Terminology by Language Independent Methods Forum Translationswissenschaft: Translation Studies and Translation Practice (19). 141-147.
45. Seljan, S., Škof Erdelija, N., Kučič, V., Dunder, I. & Pejić Bach, M. (2021). Quality Assurance in Computer-Assisted Translation in Business Environments. *Natural Language Processing for Global and Local Business*. 242-270.
46. Sreelekha, S. (2017). Statistical Vs Rule Based Machine Translation: A Case Study on Indian Language Perspective. Springer Journal of Advances in Intelligence and Soft Computing. 663-675. Doi: 10.1007/978-981-10-5520-1_60
47. Trados (bez dat.). Trados Studio Freelance. Dostupno 10.7. 2023. na: <https://www.trados.com/products/trados-studio/freelance/>
48. Trados (bez dat.). Translation Memory. Translation 101. Dostupno 15.5.2023. na: <https://www.trados.com/solutions/translation-memory/translation-101-what-is-translation-memory.html>.
49. Trados (bez dat.). What is Translation Alignment?. Dostupno 20.5.2023. na: <https://www.trados.com/solutions/translation-alignment/>.

50. Warburton, K. (2006) Terminology as a Key Driver in Business Communications, Bridging the Divide Between Authoring and Translation, *The Globalization Insider*.
51. Webb, L. E. (1998). Advantages and Disadvantages of Translation Memory: A Cost/Benefit Analysis. Monterey Institute of International Studies. Monterey, CA.
52. Zaretskaya, A., Corpas Pastor, G. & Seghiri M. (2015). Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. *Journal of Translation and Interpretation*. 8(1). 76-88.

Popis slika

Slika 1: Ručno segmentiranje	33
Slika 2: Sravnjivanje datoteka	34
Slika 3: Izvješće alata <i>Trados Studio</i>	37
Slika 4: Prikaz pogreške nejednakosti broja rečenica	42
Slika 5: Prikaz pravopisne pogreške	42
Slika 6: Prikaz usporedbe sličnosti prijevoda u alatu <i>GoTranscript</i>	51

Popis tablica

Tablica 1: Ispravci prijedloga strojnog prevođenja	38
Tablica 2: Usporedba vremena prevođenja	46
Tablica 3: Broj pogrešaka tradicionalnog prevođenja	47
Tablica 4: Pravopisne pogreška tradicionalnog prevođenja	48
Tablica 5: Gramatičke pogreške tradicionalnog prevođenja	49
Tablica 6: Odbačene gramatičke pogreške tradicionalnog prevođenja.....	50
Tablica 7: Usporedba sličnosti prijevoda	52
Tablica 8: Krivi prijevodi termina u tradicionalnom prevođenju	53
Tablica 9: Krivi prijevodi termina u prevođenju pomoću CAT alata	55

Popis grafova

Grafikon 1: Usporedba vremena prevođenja.....	46
Grafikon 2: Usporedba sličnosti prijevoda.....	52

Primjena sustava za računalno potpomognuto prevođenje u domeni obrazovanja

Sažetak

Cilj ovog rada jest analizirati mogućnosti primjene sustava za računalno potpomognuto prevođenje u domeni obrazovanja. Rad je sastavljen od dva osnovna dijela: teorijskog i praktičnog. U teorijskom prikazat će se uloga i primjena alata za računalno potpomognuto prevođenje (Computer-Assisted Translation - CAT), upravljanje tijekom procesa rada, upravljanje terminologijom i postupci kontrole kvalitete.

Rezultati istraživanja su prikazali kraće vrijeme prevođenja korištenjem CAT alata, smanjenje broja pogrešaka, dosljednost u prevođenju te snalaženje terminologijom. Osnovni nedostaci istraživanja su prevođenje istog teksta i kod tradicionalnog prevođenja i kod prevođenja pomoću CAT alata, mala količina prikupljenih resursa i mali broj ispitanika, zbog čega se ovo pilot istraživanje smatra tek inicijalnim uvidom u moguće prednosti i nedostatke korištenja CAT alata.

Ključne riječi: CAT, računalno potpomognuto prevođenje, strojno prevođenje, Trados Studio, evaluacija, analiza, informacijske znanosti, obrazovanje

Applying computer-assisted translation systems in the field of education

Summary

The aim of this paper is to analyze the potential applications of Computer-Assisted Translation (CAT) systems in the field of education. This paper is composed of two main sections: theoretical and practical. In the theoretical section, it is presented the role and usage of Computer-Assisted Translation tools, workflow management, terminology management, and quality control procedures.

The research results have shown shorter time in translation when using CAT tools, a reduction in the number of errors, consistency in translation, and proficiency in using terminology. The primary limitations of the study include: translation of the same text in both traditional translation and CAT-assisted translation, limited amount of gathered resources and a small number of participants. Therefore, this pilot study is considered an initial exploration of the potential advantages and disadvantages of using CAT tools.

Key words: CAT, Computer-Assisted Translation, machine translation, Trados Studio, evaluation, analysis, information sciences, education.