

Filozofski fakultet
Odsjek za Informacijske i komunikacijske znanosti
Ak. god. 2016./2017.

Nikolina Škof

**Utjecaj digitalnih resursa i alata na kvalitetu prijevoda u
računalno potpomognutom prevođenju**

Diplomski rad
prof. dr. sc. Sanja Seljan

Zagreb, 2017.

Zahvaljujem svim mentorima i kolegama prevoditeljske tvrtke Ciklopea na ukazanom povjerenju, pruženom znanju i tehničkoj podršci prilikom pisanja ovog diplomskog rada.

Zahvaljujem i mentorici prof. dr. sc. Sanji Seljan na motivaciji i poticanju interesa za područje jezičnih tehnologija tijekom cijelog studija.

Sadržaj

SAŽETAK	4
ABSTRACT.....	4
1. UVOD.....	5
2. RAZVOJ STROJNOG PREVOĐENJA.....	6
3. RAČUNALNO POTPOMOĞNUTO PREVOĐENJE.....	8
3.1. PREVODITELJSKE RADNE STANICE.....	8
3.2. RESURSI I ALATI U PREVOĐENJU	9
3.3. LOKALIZACIJA.....	10
3.4. PRIJEVODNE MEMORIJE	12
3.5. SRAVNJIVANJE	13
3.6. TERMINOLOŠKE BAZE.....	15
4. OSIGURANJE KVALITETE PRIJEVODA.....	17
5. PRAKTIČNI DIO.....	20
5.1. UVOD	20
5.2. SRAVNJIVANJE TEKSTOVA	20
5.3. EKSTRAKCIJA TERMINOLOGIJE	23
5.4. ANALIZA KVALITETE PRIJEVODA	26
5.4.1. Tipovi grešaka.....	26
5.4.2. Stupanj pogreške (engl. Level of mistake)	28
5.4.3. Alati.....	28
5.4.4. Analiza rezultata QA alata.....	42
5.4.5. Dodatne mogućnosti alata	49
5.5. EVALUACIJA KVALITETE PRIJEVODA	57
5.5.1. Višedimenzionalna metrika za evaluaciju kvalitete	57
5.5.2. Evaluacija kvalitete.....	58
5.5.3. Rasprava	61
6. ZAKLJUČAK.....	62
7. LITERATURA	63
POPIS SLIKA	67
POPIS TABLICA	68

Sažetak

Ovaj diplomski rad proučava utjecaj digitalnih resursa i alata na kvalitetu prijevoda u računalno potpomognutom prevođenju (engl. *Computer-Aided Translation* – CAT). U radu je prikazan pregled najvažnijih definicija iz spomenutog područja, a kvaliteta je analizirana u okviru procesa osiguranja kvalitete prijevoda (engl. *Quality Assurance* – QA). Sama provjera kvalitete je provedena na dvojezičnom englesko-hrvatskom tekstu, dobivenom postupkom sravnjivanja korpusa. U radu je korišteno nekoliko alata za računalno potpomognuto prevođenje u više različitih zadataka: sravnjivanju dokumenata, ekstrakciji terminologije te analizi kvalitete prijevoda. Njihova će usporedba pokazati koliko se kvaliteta prijevoda pospješuje korištenjem svih mogućnosti koje ti alati nude.

Ključne riječi: *kvaliteta prijevoda, računalno potpomognuto prevođenje (CAT), osiguranje kvalitete prijevoda (QA), sravnjivanje, ekstrakcija terminologije, alati za računalno potpomognuto prevođenje*

Abstract

This master's thesis analyses the impact of digital resources and tools on translation quality in computer-aided translation (CAT). Thesis presents the most important definitions from the mentioned area, and quality is analysed through the quality assurance process (QA). The quality is checked on bilingual english-croatian texts, obtained by alignment of corpus of texts written in those languages. Several CAT tools are used in different tasks: alignment, terminology extraction and analyse of translation quality. Their comparison will show how much the translation quality increases by using all the features they offer.

Key words: *translation quality, computer-aided translation (CAT), alignment, terminology extraction, CAT tools*

1. Uvod

Razvoj računalne tehnologije je uvelike utjecao na jezičnu industriju pa se tako prevođenju u današnje vrijeme pristupa s nekih potpuno drugačijih stajališta u odnosu na ona koja su postojala prije nekoliko desetljeća. Korištenje računala i u ovom području donosi mnoge prednosti, no i povećava zahtjeve u vidu količine posla, dodatnog obrazovanja te nabavke novih alata i resursa. Osim već postojećih prevoditeljskih poslova koji su usmjereni samo na klijente i prevoditelje, sve više rastu i dodatni zadaci u sklopu upravljanja projektom, a posao se više ne usmjerava samo na prevoditelje, već i na ljude drugih poslovnih profila, kao što su voditelji projekata, terminolozi, kontrolori kvalitete i informacijski stručnjaci. Postupak prevođenja je porastao za još nekoliko djelatnosti te zahtijeva dodatnu naobrazbu vezanu uz razne tehnološke zadatke. Kako bi im računalo uistinu služilo kao pomoćni alat u izvršavanju zadatka i kako bi iskoristili sve prednosti tehnologije, od velike je važnosti uložiti u znanje i edukaciju ljudi zaposlenih u određenoj jezičnoj industriji. Tek tada je moguće govoriti o svim prednostima koje donosi korištenje alata. Ovaj diplomski rad dat će nekoliko primjera korištenja resursa i alata u jezičnoj industriji te ukazati na njihov utjecaj na proces osiguranja kvalitete prijevoda.

U prvom dijelu rada prikazana je teorijska osnova u okviru računalno potpomognutog prevođenja te definicije ključne za razumijevanje istraživanja provedenog u drugom dijelu rada. Istraživanje je podijeljeno u nekoliko zadataka. Prvi zadatak je sravnjivanje dokumenata, a zatim slijede ekstrakcija terminologije, analiza kvalitete prijevoda i evaluacija rezultata dobivenih analizom. Rezultati dobiveni analizom alata poslužit će za računanje ocjene kvalitete prijevoda. Važnost osiguranja kvalitete prijevoda povećava se zajedno s rastućom potrebom za prijevodima, koja je uslijedila pojavom alata za računalno potpomognuto prevođenje i njihova neprestanog usavršavanja. Cilj praktičnog dijela ovog rada je prikazati mogućnosti korištenja resursa i alata za provjeru kvalitete u svrhu njezina poboljšanja. Analiza alata za provjeru kvalitete sastojat će se od istraživanja njihove mogućnosti prilagodbe zahtjevima projekta (kao što su terminološka baza, lista termina koji se ne prevode i sl.), a rezultati analize dat će podatak o tome koliko se njihovim korištenjem pospješuje kvaliteta prijevoda.

2. Razvoj strojnog prevođenja

Prema (Hutchins, 1986), ideja za sistematizacijom prevođenja povezuje se s R. Descartesovom idejom o „univerzalnom jeziku” iz 17. stoljeća. „Univerzalni jezik” zasniva se na jedinstvenom simbolu ili kodu, leksički različito reprezentiranom među jezicima.

Iako u to vrijeme samo ideja, tri stoljeća kasnije ona se uspijeva realizirati u okviru novog pristupa– strojnom prevođenju. Prve praktične primjene u znanosti o strojnom prevođenju povezuju se s patentima dvojice istraživača tridesetih godina prošlog stoljeća. George Artsrouni je 1933. patentirao uređaj za prevođenje koji je nazvao „Mechanical Brain”, a temeljio se na sistemu pohrane na papirnatoy vrpici. Sličnu ideju je realizirao i Petr Petrovich Smirnov-Troyanskii nekoliko mjeseci kasnije u obliku svog uređaja za prevođenje koji se temeljio na automatiziranom rječniku i omogućavao prevođenje na više jezika istovremeno. Princip prevođenja sastojao se u analizi i pretvorbi jezika izvornika u logičku formu (npr. nominativ za imenice), zatim pretvorbe logičke forme izvornika u logičku formu ciljnog jezika (uz pomoć uređaja) te ponovno pretvorba logičke forme ciljnog jezika u riječ ili rečenicu tog jezika. Unatoč tome što je samo drugi stadij procesa prevođenja automatiziran, Troyanskii je imao ideju da se automatiziraju i prva i zadnja etapa (pretvorba jezika u logičku formu i obrnuto), što će nastupiti tek godinama kasnije.¹

Značajnije doba za strojno prevođenje nastupa 50-ih godina prošlog stoljeća . 1949. godine matematičar Warren Weaver objavljuje memorandum o prevođenju u kojem ističe nekoliko važnih točaka: problem višestrukog značenja u jeziku, logičku osnovu jezika, korištenje aplikacija komunikacijske teorije i kriptografije te univerzalnost jezika.² Nakon Weaverova memoranduma slijedi niz godina razvoja strojnog prevođenja u obliku ideja, održavanja konferencija i prvih sustava za strojno prevođenje. Najveću pažnju je privukla 1954. godine javna demonstracija strojnog prevođenja u sklopu IBM-ovog projekta na Sveučilištu u Georgetownu, gdje je 49 rečenica ruskog jezika prevedeno na engleski uz pomoć rječnika od 200 riječi i 6 gramatičkih pravila.³ Ova javna demonstracija je rezultirala povećanim interesom za područje strojnog prevođenja sredinom prošlog stoljeća, no desetak godina

¹ Hutchins, John. The precursors and the pioneers. U: Machine translation: past, present, future. New York: Halsted Press, 1986. Str. 2-3. URL: <http://www.hutchinsweb.me.uk/PPF-TOC.htm> [10.9.2017.]

² Hutchins, John. The precursors and the pioneers. U: Machine translation: past, present, future. New York: Halsted Press, 1986. Str. 13. URL: <http://www.hutchinsweb.me.uk/PPF-TOC.htm> [10.9.2017.]

³ Dovedan, Zdravko; Seljan, Sanja; Vučković, Kristina. Strojno prevođenje kao pomoć u procesu komunikacije. Str. 283-291. Informatologia 35 (4), 2002. Str. 3. URL: <http://dzs.ffzg.unizg.hr/text/spuk.pdf> [10.9.2017.]

kasnije nastupa tihi period, kao rezultat objave ALPAC-ova izvješća u kojem je iznešena kritika na sustave za strojno prevođenje.⁴ Unatoč tome što je izvještaj ALPAC-a doveo do smanjenog interesa za strojnim prevođenjem, donio je i nekoliko pozitivnih pomaka u razmišljanjima i smjerovima korištenja računala u prevođenju. Očekivanja strojnog prevođenja su se smanjila i postala realističnija, a fokus u prevođenju više nije bio samo na automatskoj proizvodnji prijevoda, nego se počinje ulagati u alate i računalna pomagala koja bi pospješila rezultate računalnog prevođenja, imajući na umu problematiku značenja i razumijevanja prirodnog jezika. Osamdesetih godina prošlog stoljeća postaje sasvim jasno da proizvodnja automatskog prijevoda visoke kvalitete nije cilj koji je moguće dostići strojnim prevođenjem (. Javlja se ideja korištenja sustava za strojno prevođenje unutar kontroliranog jezika, tj. podjezika prirodnog jezika koji se sastoji od limitiranih leksičkih i sintaktičkih struktura i odnosi se na usko specijalizirano područje. Primjer jednog od prvih sustava za strojno prevođenje na temelju podjezika je kanadski sustav Météo, koji prevodi vremenske prognoze s engleskog na francuski jezik. Primjena strojnog prevođenja u ovakvim slučajevima može ispuniti očekivanja, no svejedno se, paralelno uz sustave za automatsko strojno prevođenje, počinje raditi na alatima i resursima koji će služiti samo kao pomoć u prevođenju i u dostupu informacija.⁵

⁴ Hutchins, John. Expectations and criticisms. U: Machine translation: past, present, future. New York: Halsted Press, 1986. Str. 15. URL: <http://www.hutchinsweb.me.uk/PPF-8.pdf> [10.9.2017.]

⁵ Somers, Harold. Computers and Translation: A Translator's Guide. Philadelphia. John Benjamins Publishing Company, 2003. Str. 6-7

3. Računalno potpomognuto prevođenje

Za razliku od strojnog prevođenja kod kojeg je u središtu sustav za strojno prevođenje i automatska proizvodnja prijevoda, u računalno potpomognutom prevođenju u središtu je prevoditelj, koji se računalom i računalnim resursima služi za pomoć u postupku prevođenja.⁶ Prvim računalnim resursima korištenim za pomoć čovjeku u procesu prevođenja smatraju se automatski rječnici i terminološke baze. Kasniji razvoj resursa i alata dovodi do pojave prijevodnih memorija, kao jedne od najznačajnijih komponenata uključenih u računalno potpomognuto prevođenje. Osim rječnika, terminoloških baza i prijevodnih memorija, u alate i resurse ubrajaju se i provjernici pravopisa i gramatike, alati za sravnjivanje, sustavi za upravljanje terminologijom, alati za ekstrakciju terminologije te oni za osiguranje kvalitete prijevoda. Većina CAT alata u sebi sadrži barem nekoliko nabrojanih resursa, iako oni mogu biti i samostalne aplikacije. Zbog sve većeg korištenja alata i tehnologije općenito, javlja se novo radno okruženje u okviru računalno potpomognutog prevođenja naziva se prevoditeljska radna stanica (engl. *translator's workbench* ili *translator's workstation*).⁷

3.1. Prevoditeljske radne stanice

U ranim 90-im godinama prošlog stoljeća na tržištu se pojavljuje nekoliko komercijalnih prevoditeljskih radnih stanica za profesionalne prevoditelje. Tvrtnica IBM Corporation je predstavila TranslationManager/2, a njemačka tvrtka STAR AG je predstavila Transit sustav. U isto vrijeme pojavljuje se i Translator's Workbench tvrtke Trados, koja je prva u svoju prevoditeljsku radnu stanicu uklopila prijevodnu memoriju i sravnjivanje.⁸ Prevoditeljske radne stanice na jednom mjestu okupljaju prevoditelje, voditelje projekata, terminologe, kontrolore kvalitete, programere i informacijske stručnjake. . Osim ljudi različitih profila, tu se nalaze i alati uključeni u različite procese obrade teksta i jezika, kao što su alati za optičko prepoznavanje znakova (OCR), pretvorbu formata, ali i segmentaciju, poravnanje dokumenata, upravljanje terminologijom, prijevodnim memorijama i kvalitetom. Uzimajući u obzir velik broj alata i resursa koji se neprestano pojavljuju na tržištu, ne treba uopće čuditi koliki značaj za jezičnu industriju je imala pojava prevoditeljskih radnih stanica.

⁶ Hutchins, John. Current commercial machine translation systems and computer-based translation tools: system types and their uses. 2005. URL: <http://www.hutchinsweb.me.uk/IJT-2005.pdf> [10.9.2017.] Str. 12

⁷ Ibid.

⁸ Hutchins, John. The origins of the translator's workstation. 1998. University of East Anglia, Norwich, UK. URL: <https://pdfs.semanticscholar.org/0bb2/fe15d30b1243f2fc90bed2ecc5a23d57ccfc.pdf> [10.9.2017.] Str. 15

3.2. Resursi i alati u prevođenju

Alcina (2008)⁹ daje klasifikaciju pet glavnih komponenata tehnologije u prevođenju:

1. Računalna oprema,
2. Alati za komunikaciju i dokumentaciju,
3. Uređivači teksta i materijala za objavljivanje,
4. Jezični alati i resursi,
5. Alati za prevođenje.

Prvu skupinu čine fizičke komponente računala potrebne prevoditelju za rad, operacijski sustav, softveri za održavanje sustava (npr. antivirusni program), uređaji za čitanje i pisanje podataka (npr. cd-rom), monitor, tipkovnica, miš, printer, skener i ostalo.¹⁰

U komunikacijske i dokumentacijske alate ubrajaju se alati i resursi koji služe za interakciju s klijentima, ostalim sudionicima projekta, terminolozima ili drugim ljudima uključenima u projekt.¹¹

Uređivači teksta su alati koji služe za obradu (unošenje, izmjenu ili reviziju) teksta, ali i uređivači html-a i/ili xml-a i aplikacija za dizajn web stranica. Osim svojih osnovnih funkcija, današnji alati nude razne mogućnosti kao što su uključivanje provjere pravopisa, gramatike ili stila, upravljanje unošenjem izmjena u dokument, kreiranje predložaka itd.¹²

Jezični alati i resursi služe za prikupljanje i organizaciju lingvističkih jedinica (riječi, termina, tekstova) s dodatnim lingvističkim informacijama o tim jedinicama (gramatička kategorija riječi, žanr tekstova i sl.). Autor naglašava razliku između alata i resursa, pri čemu alati služe za upravljanje vlastitim dokumentima i informacijama, a resursi su već gotovi, prikupljeni tekstualni podaci. Ova distinkcija između alata i resursa može se primijeniti i na praktični dio ovog rada. Alati korišteni u radu pomažu u obavljanju određenog zadatka (npr. za ekstrakciju

⁹ Alcina, A., Translation Technologies: Scope, tools and Resources, International Journal on Translation Studies. Španjolska. 2008. URL: <http://repositori.uji.es/xmlui/bitstream/handle/10234/18016/28696.pdf> [5.9.2017.] Str. 23

¹⁰ Ibid.

¹¹ Ibid.

¹² Ibid. Str. 24

terminologije je korišten alat memoQ), a resurs čini gotov proizvod tog zadatka (npr. terminološka baza).¹³

Zadnja komponenta tehnologije u prevođenju su alati. U njih se ubrajaju ujedno i programi za računalno potpomognuto prevođenje (zajedno sa sustavima za prijevodne memorije, upravljanje terminologijom i programima za obradu teksta) te programi za strojno prevođenje.¹⁴

3.3. Lokalizacija

Upotreba računala u prevođenju ne zahtijeva samo alate i softvere, već i one koji će podržavati različito jezično okruženje, primjerice računalno prepoznavanje znakova (engl. *optical character recognition* – OCR) velikog broja jezika, provjernike pravopisa različitih jezika, korisničko sučelje prilagođeno različitim jezicima i sl. Jedno od najbrže rastućeg područja primjene računala u prevođenju je lokalizacija softvera. 1990. godine osnovana je udruga za standarde u lokalizacijskoj industriji (engl. *Localisation Industry Standards Association* – LISA). Cilj je udruge promoviranje lokalizacijske i internacionalizacijske industrije te omogućavanje razmjene informacija među kompanijama o razvoju procesa, alata, tehnologija i poslovnih modela povezanih s lokalizacijom i internacionalizacijom.¹⁵

Postupak internacionalizacije prethodi lokalizaciji i odnosi se na generalizaciju proizvoda koji će podržavati više jezika, bez potrebe za redizajnom programa.¹⁶ Lokalizacija obuhvaća pretvaranje proizvoda lingvistički i kulturalno prilagođenog ciljnom tržištu, s ciljem korištenja proizvoda na vlastitom jeziku bez ikakvih teškoća.¹⁷ Unutar prevoditeljske radne stanice lokalizacijski poslovi odnose se prvenstveno na prevođenje, ali i ostale postupke jezične prilagodbe, kao što su prilagodba pisma softvera, elemenata tipkovnice, fontova te raspored grafičkih i tekstualnih elemenata, formatiranje datuma, vremena, mjernih i novčanih jedinica.

¹³ Alcina, A., Translation Technologies: Scope, tools and Resources, International Journal on Translation Studies. Španjolska. 2008. URL: <http://repositori.uji.es/xmlui/bitstream/handle/10234/18016/28696.pdf> [5.9.2017.] Str. 25

¹⁴ Ibid. Str. 26

¹⁵ Esselink, Bert. A Practical Guide to Localization. John Benjamins Publishing Company. Amsterdam/Philadelphia. 2000. Str. 8

¹⁶ Ibid. Str. 2

¹⁷ Esselink, Bert. The Evolution of Localization. Solution Architect, Lionbridge. 2000. URL: http://www.intercultural.urv.cat/media/upload/domain_317/arxiu/Technology/Esselink_Evolution.pdf [10.9.2017.] Str. 1

Osim toga, lokalizacijski poslovi obuhvaćaju i višejezično upravljanje projektima, izgradnju i testiranje softvera i mrežne pomoći, konvertiranje prevedene dokumentacije u druge formate, sravnjivanje i upravljanje segmenata u prijevodnim memorijama, izgradnju višejezične podrške za proizvode i savjetovanje u vezi s prijevodnim strategijama. vrsta pisma softvera, elementi tipkovnice, fontovi te raspored grafičkih i tekstualnih elemenata.¹⁸

Za potrebe lokalizacije softvera inženjeri dobivaju kopiju softverskog razvojnog okruženja, izdvajaju datoteke s tekstom za prevođenje i pripremaju ga te podupiru prevoditelje tijekom njihovog rada. Gotove je prijevode potrebno ponovno spojiti i ubaciti u razvojno okruženje te kompilirati lokaliziranu kopiju softvera i testirati ga, kako bi se uklonile postojeće greške.

U lokalizaciji on-line sustava za pomoć koristi se sličan pristup, izvorni dokumenti (najčešće RTF ili HTML datoteke) se prevode, kompiliraju i testiraju.¹⁹ Lokalizacijski alati koji se koriste u lokalizaciji softvera, kao što su SDL Passolo, Alchemy Catalyst, Language localizator, Multilizer, razlikuju se od ostalih CAT alata po tome što ne narušavaju programski kod, podržavaju niz softverskih formata i izgledom su prilagođeni lokalizacijskom okruženju. Bez obzira na to radi li se o alatu posebno namijenjenom lokalizaciji ili o nekom drugom CAT alatu, resursi kao što su prijevodne memorije i terminološke baze uvelike pomažu u održavanju konzistentnosti i ubrzavaju proces prevođenja, primjerice između različitih verzija softvera koje se ažuriraju.

Osim lokalizacije softvera, važnost korištenja nabrojanih resursa postoji i u lokalizaciji web stranica, čiji se sadržaj često mijenja i nadopunjuje. Prijevodna memorija omogućuje automatsko unošenje starog prijevoda u segmente koji nisu mijenjani, tako da ostaju samo oni koje je potrebno po prvi puta prevesti. Najveća je uloga terminološke baze u očuvanju konzistentnosti termina i kompletnog prijevoda unutar različitih verzija softvera, web stranica i ostalih lokalizacijskih projekata. U današnje doba kad potrebe za jezičnom prilagodbom u internetskom i računalnom okruženju uvelike rastu, poželjno je razvijati alate i resurse poput spomenutih prijevodnih memorija i terminoloških baza.

¹⁸ Škrbina, Boris. Diplomski rad, Računalni alati u izradi prijevodne memorije, Zagreb, Filozofski fakultet, 2014. URL: <http://darhiv.ffzg.unizg.hr/id/eprint/4583> [10.9.2017.] Str. 33

¹⁹ Esselink, Bert. The Evolution of Localization. Solution Architect, Lionbridge. 2000. URL: http://www.intercultural.urv.cat/media/upload/domain_317/arxiu/Technology/Esselink_Evolution.pdf [10.9.2017.] Str. 6

3.4. Prijevodne memorije

Sustavi za prijevodne memorije na tržištu postoje već tridesetak godina, a od njihove pojave pa sve do danas glavni cilj im se nije promijenio: pregledavanje i primjena starih prijevoda u istim ili sličnim segmentima prijevoda.²⁰ prijevodne memorije definira kao baze podataka koje se sastoje od parova segmenata, a segment se ovdje odnosi na jedinicu izvornika i prijevoda, koja može biti duljine jedne riječi ili fraze, ali i cijele rečenice ili (rjeđe) odlomka.

Osim u okviru lokalizacije softvera i lokalizacijskim alatima o kojima je do sad bilo riječi, prednosti korištenja prijevodnih memorija u vidu uštede vremena i očuvanja konzistentnosti, važne su i za sve druge tipove projekata. U radu za istog klijenta, nije rijetka pojava već prethodno prevedenih segmenata, a zadatak više nije njihovo ponovno prevođenje, već pregledavanje automatskog unosa iz memorije, njegovo potvrđivanje ili uređivanje. Ovisno o alatu i njegovim mogućnostima, nekoliko je različitih stupnjeva podudaranja segmenata iz memorije.

Uobičajena je klasifikacija na one 100 %-tne podudarnosti s memorijom (engl. *exact match*), djelomične podudarnosti s memorijom (engl. *fuzzy match*) i 101 %-tne podudarnosti s memorijom (to su oni segmenti koji se međusobno 100 % podudaraju, ali se podudaraju i u segmentima koji im prethode i slijede nakon njih, engl. *context match*).

Sadržaj prijevodne memorije su prijevodne jedinice (engl. *translation units*) koje se sastoje od segmenta izvornika i segmenta prijevoda.²¹ naglašava dinamičnu prirodu prijevodnih memorija i definira tri moguća načina njezinog rasta: prilikom prevođenja (potvrđivanjem segmenata oni se automatski spremaju u prijevodnu memoriju dodijeljenu projektu), importiranje nove prijevodne memorije u postojeću (ona može dolaziti iz bilo kojeg drugog alata, no mora biti u standardnom *.tmx formatu), sravnjivanjem tekstova izvornika i prijevoda (ručno kreiranje prijevodne memorije na temelju postojećih prevedenih tekstova).

Unatoč tome što sustavi za prijevodne memorije postoje već dugi niz godina i što im se glavni cilj nije izmijenio, potrebe za razvijanjem novih tehnologija u radu s prijevodnim memorijama konstantno rastu.

²⁰ Zerfass, Angelika. Evaluating Translation Memory Systems. 2002. <http://www.mt-archive.info/LREC-2002-Zerfass.pdf> [10.9.2017.] Str. 1

²¹ Reinke, Uwe. State of the Art in Translation Memory Technology. Cologne University of Applied Sciences. 2013. URL: <http://www.blogs.uni-mainz.de/fb06-tc3/files/2015/11/25-145-3-PB.pdf> [05.9.2017.] Str. 7

Prema (Reinke, 2013)²², tipičan sustav za prijevodnu memoriju sastoji se od višejezičnog uređivača teksta, programa za upravljanje terminologijom, funkcije automatskog prepoznavanja termina, alata za konkordancije, računanje statistike i alata za sravnjivanje. Dodatna je mogućnost sustava za prijevodne memorije i opcija pretprijevoda (engl. *pre-translate*), koja se sastoji od pregledavanja prijevodne memorije i automatskog unošenja podudaranja prema željenom postotku (primjerice, alati nude mogućnost namještanja automatskog unošenja samo na *exact matches*, ali i segmente koji se podudaraju u minimalno 50 % ili u nekom drugom postotku). Slična je funkciji pretprijevoda i funkcija „*fragment assembly*”, koja se sastoji od sastavljanja prijevoda od dijelova koji se podudaraju s memorijom. Dodatno je uređivanje sastavljenih segmenata nužno te je, ovisno o situaciji, potrebno procijeniti isplati li se ovakav pristup u određenom slučaju i hoće li uređivanje segmenata zahtijevati više vremena od ponovnog prevođenja.

S obzirom na to da se radi o novoj tehnologiji čije se mogućnosti još uvijek razvijaju i pospješuju, na umu treba imati sve moguće poteškoće koje mogu nastati prilikom spajanja dijelova segmenata iz memorije. Jedan od vodećih prevoditeljskih softvera i alata za računalno potpomognuto prevođenje, SDL Trados, ove je godine predstavio upLift tehnologiju, koja omogućava tzv. *Fragment Recall*, odnosno pozivanje subsegmenata (fragmenata) iz memorije prilikom prevođenja.²³ Ono što se također izmijenilo od pojave sustava za prijevodne memorije je i radni proces, odnosno organizacija podataka, suradnja i interakcija ljudi uključenih u prevoditeljskih projekt.²⁴ Pojava serverskih prijevodnih memorija omogućuje rad više korisnika na istoj prijevodnoj memoriji, koja može biti višejezična i koja je spremljena na serveru klijenta i zahtijeva prijavu kako bi joj se moglo pristupiti.

3.5. Sravnjivanje

Jedan od spomenutih načina dobivanja prijevodne memorije je bio sravnjivanjem paralelnih tekstova izvornika i prijevoda. Sravnjivanje se definira kao postupak paralelne segmentacije dvaju tekstova, obično u manje, logične jedinice, kao što su rečenice, tako da

²² Reinke, Uwe. State of the Art in Translation Memory Technology. Cologne University of Applied Sciences. 2013. URL: <http://www.blogs.uni-mainz.de/fb06-tc3/files/2015/11/25-145-3-PB.pdf> [05.9.2017.] Str. 3

²³ SDL Trados. upLIFT technology - the difference is clear to see. Web page. URL: <http://www.sdltrados.com/products/trados-studio/uplift/> [10.9.2017.]

²⁴ Reinke, Uwe. State of the Art in Translation Memory Technology. Cologne University of Applied Sciences. 2013. URL: <http://www.blogs.uni-mainz.de/fb06-tc3/files/2015/11/25-145-3-PB.pdf> [05.9.2017.] Str. 6

svaki n-ti segment prijevoda odgovara n-tom segmentu izvornika.²⁵ Pojmovi logičkih jedinica i rečenica variraju od prijevoda do prijevoda i od jezika do jezika. Različiti alati za automatsko sravnjivanje nude različite mogućnosti definiranja pravila segmentacije.²⁶ Najčešće su to interpunkcijski znakovi iza kojih slijedi veliko slovo, no već ovdje možemo naići na iznimke kao što su kratice, koje završavaju točkom, a iza njih može slijediti, primjerice, naziv. Alat će to prepoznati kao završetak rečenice i segment razlomiti na dva dijela, dok u paralelnom segmentu prijevoda to ne mora biti tako, jer kratica možda uopće ne postoji ili taj jezik ne zahtijeva pisanje točke na njezinom kraju. Drugu poteškoću alatima za automatsko sravnjivanje predstavljaju i datumi ili samo godine, koji u hrvatskom uvijek završavaju točkom, a u nekim drugim jezicima to nije slučaj (primjerice u engleskom ili njemačkom jeziku). Zbog problema kao što su ovi navedeni, ljudski je pregled automatski sravnatih segmenata nužan.

Simard i Plamondon (1998)²⁷ nabrajaju dvije glavne poteškoće s kojima se susreću alati za automatsko sravnjivanje: robusnost i točnost. Robusnost se odnosi na razlike u rasporedu riječi u rečenicama izvornika i prijevoda, propuste ili inverzije. Poteškoće s točnošću rezultata sravnjivanja postoje u slučajevima kad je (čak i čovjeku) teško odlučiti o pripadajućim segmentima izvornika i prijevoda. Iz tog razloga se razvijaju modeli koji se ne temelje samo na definiranju pravila segmentacije prema pojavama nekih tipičnih jedinica u jeziku, kao što su interpunkcijski znakovi, već u obzir uzimaju i same riječi i njihova značenja. Primjerice, postoje modeli koji se temelje na tzv. srodnim riječima (engl. *cognates*) u jezicima, tj. onim riječima koje su korijenski i značenjski slične (npr. parlament (hr) – parliament (en) – parlement (fr)). Prvi je takav model predstavio Simard 1992., a od tad se nastavljaju razvijati modeli temeljeni na istom principu, no poboljšani u smislu da uzimaju u obzir i kontekst u rješavanju problema lažnih prijatelja u jezicima (riječi istog korijena, ali različitog značenja, engl. *false friends*), poziciju riječi u dijelovima teksta ili segmentima, njihovu dužinu i slično.

²⁵ Simard, Michel; Plamondon, Pierre. Bilingual sentence alignment: balancing robustness and accuracy. Laval (Quebec), Canada. 1998. Str 1.

²⁶ Seljan, Sanja; Gašpar, Angelina; Pavuna, Damir. Sentence Alignment as the Basis for Translation Memory Database. U: 1. međunarodna znanstvena konferencija "The Future of Information Sciences: INFUTURE2007 – Digital Information and Heritage", 2007. Zagreb.

²⁷ Simard, Michel; Plamondon, Pierre. Bilingual sentence alignment: balancing robustness and accuracy. Laval (Quebec), Canada. 1998. Str. 1

Bez obzira na to koji je pristup i način sravnjivanja u alatu korišten, njegov je cilj uvijek stvoriti što bolji i točniji resurs (prijevodnu memoriju) koji će zahtijevati što manju potrebu za ljudskom intervencijom.

3.6. Terminološke baze

Važan resurs u radu s alatima za računalno potpomognuto prevođenje su i termini, spremljeni u obliku dvojezične ili višejezične terminološke baze. Hartley (2009)²⁸ definira termine kao leksičke jedinice koje se odnose na posebno specijaliziranu domenu, a terminologiju kao proces njihova identificiranja, organiziranja i prezentiranja korisniku, ali i rezultat tih procesa – skup (često višerječnih) izraza iz specifične domene. Autor ističe važnost upotrebe konzistentne i ispravne terminologije u prevođenju, s posebnim naglaskom na današnje doba lokalizacije i globalizacije te timskog rada na projektima, što dodatno povećava mogućnost pogrešne i nedosljedne upotrebe termina. U radu na zahtjevnijim i opsežnijim prevoditeljskim projektima specijalizirane tematike, može se pojaviti potreba za uključivanjem terminologa, tj. osobe koja će upravljati terminologijom u cijelom projektu i ažurirati ju. Sustavi za upravljanje terminologijom mogu biti samostalne aplikacije ili dio nekog CAT alata. Za razliku od obične terminološke baze, omogućavaju naprednije upravljanje, pristup i prilagodbu termina specifičnoj domeni. Prevoditeljske radne stanice najčešće imaju integrirane terminološke baze u CAT alatima i to one koje serverski dijele s klijentom na projektu ili koje su dobile u nekom od formata datoteke koje terminološke baze podržavaju (npr. dvojezična ili višejezična Excel datoteka). Osim integriranih aplikacija, neki CAT alati nude zasebne aplikacije koje predstavljaju sustave za upravljanje terminologijom, kao što je Trados MultiTerm tvrtke SDL. U memoQ alatu to je QTerm sustav za upravljanje terminologijom, dostupan kao dodatak u radu na memoQ serveru.

Do sad su bile opisane situacije u kojima je terminologija dostupna od trenutka kreiranja prevoditeljskog projekta, neovisno o njezinom formatu ili načinu pristupa (serverski ili lokalno). Osim scenarija već dobivene terminološke baze, u prevoditeljskoj radnoj stanici se može pojaviti potreba kreiranja nove terminološke baze na osnovu gotovih prijevoda. U tom se slučaju radi ručna ili automatska ekstrakcija terminologije. Ručna je ekstrakcija zasigurno točniji i bolji odabir, no često i dugotrajan i mukotrpan proces. Alati za automatsku

²⁸ Hartley, Tony. *Technology and Translation*. // *The Routledge Companion to Translation Studies*. London; New York : Routledge, 2009. URL: http://www.data.ulis.vnu.edu.vn/jspui/bitstream/123456789/2910/1/UTF-8__0415396417Translation.pdf [10.9.2017.] Str. 112

ekstrakciju terminologije najčešće se oslanjaju na frekvenciju pojavljivanja riječi ili nizova riječi u izvorniku i prijevodu te zahtijevaju potvrdu prevoditelja, lektora ili terminologa o svojoj točnosti. S obzirom na to da je i čovjeku nekad teško razlučiti radi li se u određenom slučaju o terminu ili ne, niti alati još uvijek nisu toliko precizni u ekstrakciji termina, a oslanjanje na frekvenciju pojavljivanja donosi mnoge lažne kandidate (kao što su veznici, članovi, prijedlozi).

U ovom radu će biti prikazana automatska ekstrakcija terminologije kao dio memoQ alata u praktičnom dijelu rada.

4. Osiguranje kvalitete prijevoda

Porastom automatizacije u prevoditeljskim zadacima te povećanjem obujma posla dolazi do narušavanja kvalitete samih prijevoda. Unatoč tome što korištenje CAT alata donosi brojne prednosti, njihova je upotreba uvelike povećala zahtjeve za novim zadacima, što ujedno povećava mogućnost ljudske pogreške u tom procesu. Prije isporuke prijevoda klijentu je iznimno važno provjeriti njegovu kvalitetu. Kontrola kvalitete prijevoda (engl. *Quality Control* (QC)) je završni korak u projektu te dio procesa osiguranja kvalitete prijevoda (engl. *Quality Assurance* (QA)).

Unatoč tome što se pojmovi kontrole kvalitete i osiguranja kvalitete često miješaju te se QA ponekad svodi samo na završnu provjeru kvalitete prijevoda, osiguranje kvalitete je širi pojam od kontrole kvalitete i obuhvaća postupak upravljanja kvalitetom već od primitka prevoditeljskog zadatka. QA obuhvaća pregled i/ili kreiranje izvornih dokumenata, kao i promišljanje o utjecaju njihovog sadržaja na kvalitetu te moguće poteškoće u radu s alatima, dok je kontrola kvalitete tek završna provjera kvalitete prijevoda.

U ovom radu će biti razjašnjeni postupci osiguranja kvalitete prijevoda na primjeru prijevodne memorije, s posebnim naglaskom na alate i njihove mogućnosti u tom procesu. Makoushina (2008)²⁹ u proces osiguranja kvalitete prijevoda uključuje sve dokumente i resurse unutar projekta. Ponajprije su to izvornik, prijevod te finalni dokument, ali i terminološka baza, referentni dokumenti i prijevodna memorija. Kvaliteta svakog tipa dokumenta uključenog u projekt utječe na završnu kvalitetu prijevoda, stoga je važno osigurati kvalitetu resursa korištenih u projektu. Prijevodna memorija mora sadržavati visoko kvalitetne i dosljedne prijevodne parove, a prijevodne jedinice moraju biti ispravno sravnjene u memoriji, dok se terminološka baza mora sastojati od ispravnih termina.

Govoreći o kvaliteti prijevoda, nekoliko je važnih aspekata na koje treba obratiti pozornost. Makoushina (2007)³⁰ razlikuje dva pristupa u osiguranju kvalitete prijevoda. Prvi se odnosi na gramatičku točnost prijevoda, odnosno ispravnost upotrebe interpunkcije za ciljni jezik, pravopisnu provjeru te značenjsku istovjetnost s izvornikom. Drugi je pristup onaj koji se tiče formatiranja prijevoda u skladu s originalom. To uključuje provjeru istovjetnosti forme s

²⁹ Makoushina, Julia, Kockaert, J. Hendrick. Zen and the Art of Quality Assurance. 2008. URL: <http://www.mt-archive.info/Aslib-2008-Makoushina.pdf> [10.9.2017.] Str. 3

³⁰ Makoushina, Julia. Translation Quality Assurance Tools: Current State and Future Approaches. 29.11.2007., Tomsk, Russia. URL: <http://www.mt-archive.info/Aslib-2007-Makoushina.pdf> [10.9.2017.] Str. 2

originalom (često dužina prijevoda utječe na pomicanje nekih dijelova na zasebnu stranicu, vidljivost ili nevidljivost nekog dijela teksta ili slike i slično), višestrukih razmaka među riječima i višestrukih interpunkcijskih znakova, kao i provjeru dosljednosti u prijevodu i terminologije, što predstavlja idealno područje za korištenje alata u provjeri kvalitete prijevoda.

Prvi QA alati pojavili su se krajem 90-ih godina prošlog stoljeća, što je 10 – 15 godina nakon pojave prijevodnih memorija i računalno potpomognutog prevođenja. Prve provjere kvalitete uključivale su provjeru formatiranja, terminologije i pravopisa. Glavnim razlogom potrebe za osiguranjem kvalitete prijevoda smatra se neprestan razvoj tehnologije u računalno potpomognutom prevođenju, s naglaskom na sustavima za prijevodne memorije, čiji sadržaj mora biti visoke kvalitete, ali čije korištenje također može rezultirati potvrđivanjem i onih segmenata koji nisu istovjetni s prijevodom (engl. *fuzzy match*). Neoprezom se može prihvatiti segment iz memorije koji se s izvornikom podudara u više od 90 %, a razlikuje se primjerice samo u jednom broju. Takvu je grešku okom teško detektirati, za razliku od alata koji će brzo uočiti nedosljednost u brojevima izvornika i prijevoda.

U današnje vrijeme gotovo svi CAT alati omogućuju neku vrstu provjere kvalitete. Star Transit smatra se prvim CAT alatom koji je uveo nekoliko tipova provjere kvalitete kao dio prevoditeljskog procesa. Slijede ga SDLX Translation Suite, Trados, Déjà Vu i WordFast. Za razliku od provjere kvalitete prijevoda integrirane u CAT alatu, dolazi do pojave samostalnih QA alata, koji će omogućiti više tipova provjere i veću učinkovitost, zbog fokusa na samoj kvaliteti.

Prema (Makoushina, 2007), prvi samostalni QA alat koji je bio predstavljen javnosti 2004. godine u Sankt-Peterburgu je QA Distiller, razvijen u Yamagata Europe prevoditeljskoj agenciji.³¹ Najnovija verzija alata korištena je i u praktičnom dijelu ovog rada, a nastavno na QA Distiller razvijaju se i ostali samostalni QA alati na tržištu: Error Spy, ApSIC Xbench, Verifika, Okapi CheckMate.³² Unatoč tome što je od pojave prvih provjera kvalitete prijevoda u alatima prošlo gotovo dvadeset godina, mogućnosti nekih od njih nisu se puno izmijenile od tada. U nastavku će ovog rada biti analizirana četiri QA alata i testirane njihove

³¹ Makoushina, Julia. Translation Quality Assurance Tools: Current State and Future Approaches. 29.11.2007., Tomsk, Russia. URL: <http://www.mt-archive.info/Aslib-2007-Makoushina.pdf> [10.9.2017.] Str. 4

³² Kostiainen, Tuomas. Tools quality assurance and translation memory maintenance. 25.10.2012. URL: http://www.finntranslations.com/wordpress/wp-content/uploads/2010/08/LT-2_handout_TuomasKostiainen.pdf [10.9.2017.] Str. 1-2

možnosti na primjeru prijevodne memorije dobivene sravnjivanjem prevedenih tekstova iz domene poslovnog dopisivanja.

5. Praktični dio

5.1. Uvod

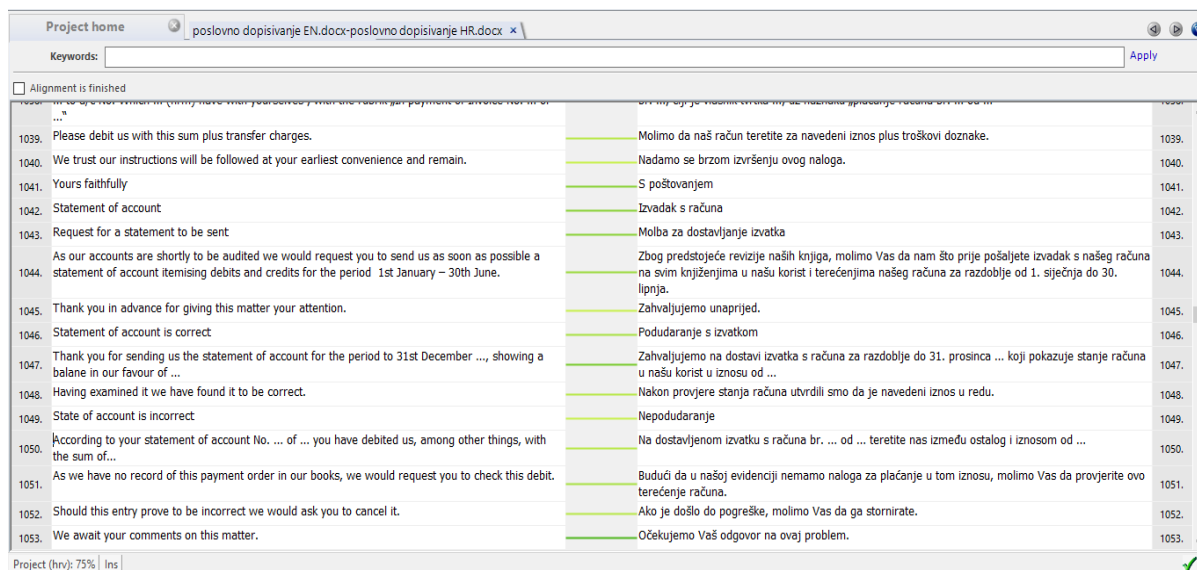
Cilj je praktičnog dijela ovog rada prikazati mogućnosti korištenja digitalnih alata i resursa u jezičnoj industriji te ukazati na njihove prednosti i moguće nedostatke. Analiza digitalnih resursa i alata će biti napravljena na prijevodima tekstova s engleskog na hrvatski jezik unutar domene poslovnog dopisivanja. Prvi je zadatak praktičnog dijela ovog rada bio sravnjivanje spomenutih tekstova izvornika i prijevoda kako bi se dobila prijevodna memorija na kojoj će se raditi analiza resursa i odabranih alata. Već je prije bilo riječi o važnosti visokokvalitetnog i dosljednog sadržaja prijevodne memorije, stoga će ovaj rad dati uvid u proces osiguranja kvalitete prijevoda na prijevodnoj memoriji dobivenoj sravnjivanjem. Za alat u kojem će se raditi sravnjivanje je odabran memoQ³³. Nakon dobivanja prijevodne memorije, u drugom je zadatku napravljena ekstrakcija termina iz memorije, koji će kasnije poslužiti kao resurs u procesu osiguranja kvalitete prijevoda. Za analizu kvalitete prijevoda su odabrana četiri alata: memoQ, Verifika, QA Distiller i Xbench. Uspješnost alata će se testirati na dobivenoj prijevodnoj memoriji, na temelju alatom pronađenih grešaka. Greške će biti podijeljene na one alatom pronađene stvarne, alatom pronađene pogrešne te one koje nisu pronađene određenim alatom. Ti će podaci poslužiti za računanje stupnja pogreške, na temelju kojeg je moguće odrediti najuspješniji od četiriju analiziranih alata za osiguranje kvalitete prijevoda. Nakon odabira najuspješnijeg alata i analize dobivenih rezultata, izračunat će se kvaliteta prijevoda metodama koje se koriste u višedimenzionalnoj metrici za evaluaciju kvalitete. Konačan će rezultat pokazati koliko se kvaliteta prijevoda poboljšava ispravljanjem grešaka koje su alati pronašli te kolika je važnost osiguranja kvalitete prijevoda u prevoditeljskom procesu.

5.2. Sravnjivanje tekstova

Postupak sravnjivanja tekstova izvornika i prijevoda počinje kreiranjem projekta u memoQ alatu. U projekt je potrebno dodati praznu prijevodnu memoriju u koju će se kasnije ubaciti sravnjeni tekstovi. Prijevodnu sam memoriju nazvala *Poslovno dopisivanje en-hr TM*. Nakon dodavanja memorije je potrebno dodati i prazan korpus u koji će se dodati par tekstova za sravnjivanje. Postupak sravnjivanja tekstova u memoQ-u sastoji se od provjeravanja automatski sravnjenih segmenata originala i prijevoda te intervencije u slučaju potrebe. Automatski sravnjeni segmenti dobivaju zelenu vezu između sebe, a potrebno ih je

³³ Verzija korištena u ovom radu: Translator pro 2015.

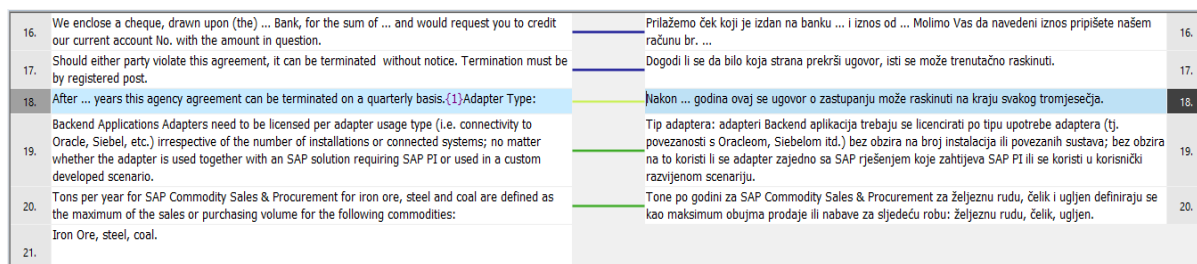
potvrditi nakon što se uvjerimo da izvorni segment uistinu odgovara prijevodnom segmentu. Na Slici 1. je prikazan rezultat pokretanja automatskog sravnjivanja tekstoiva u memoQ alatu.



Slika 1. Sravnjivanje u memoQ alatu

Unatoč naizgled dobrom rezultatu automatskog sravnjivanja dokumenata, segmente je potrebno pregledati prije potvrđivanja jer određene razlike u formatiranju izvornika i prijevoda mogu utjecati na pogrešnu segmentaciju pa tako i krivo sravnjene rečenice. Potvrđeni segmenti dobivaju plavu vezu.

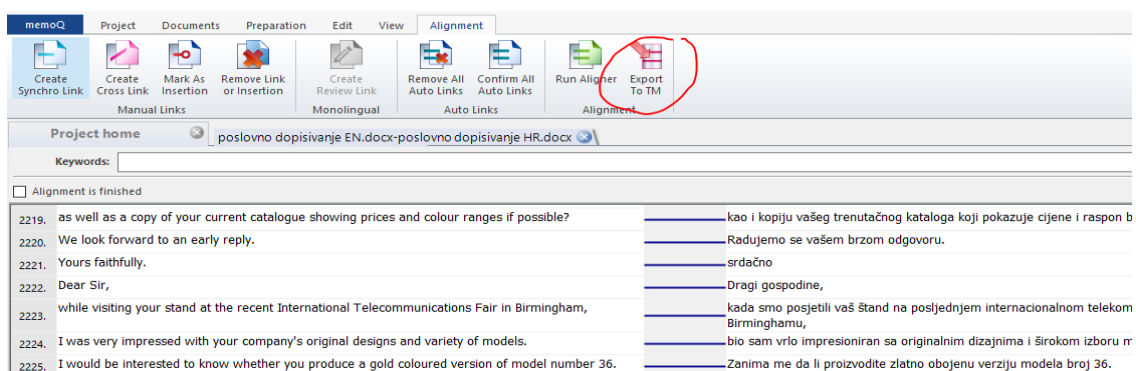
Slika 2. prikazuje primjer u kojem su vidljive dvije greške u sravnjivanju. U prvom slučaju se radi o dijelu iz osamnaestog segmenta izvornika (*Adapter Type*) koji se nalazi u devetnaestom segmentu prijevoda, što je ujedno i točna segmentacija, stoga je potrebno taj dio ubaciti u devetnaesti segment izvornika. To je moguće napraviti opcijama „Razdvoji” (engl. *Split*) i „Spoji” (engl. *Join*) ili jednostavno kopiranjem i lijepljenjem tog dijela na odgovarajuće mjesto. Druga greška, također vidljiva na Slici 2., je razlomljeni dvadeseti segment izvornika, još jednom na dijelu s dvotočkom (... for the following commodities: *Iron Ore, steel, coal.*). Kako bismo ispravili grešku, potrebno je spojiti 20. i 21. segment izvornika, kako bi konačni segment odgovarao prijevodu.



Slika 2. Greške u automatskom sravnjivanju

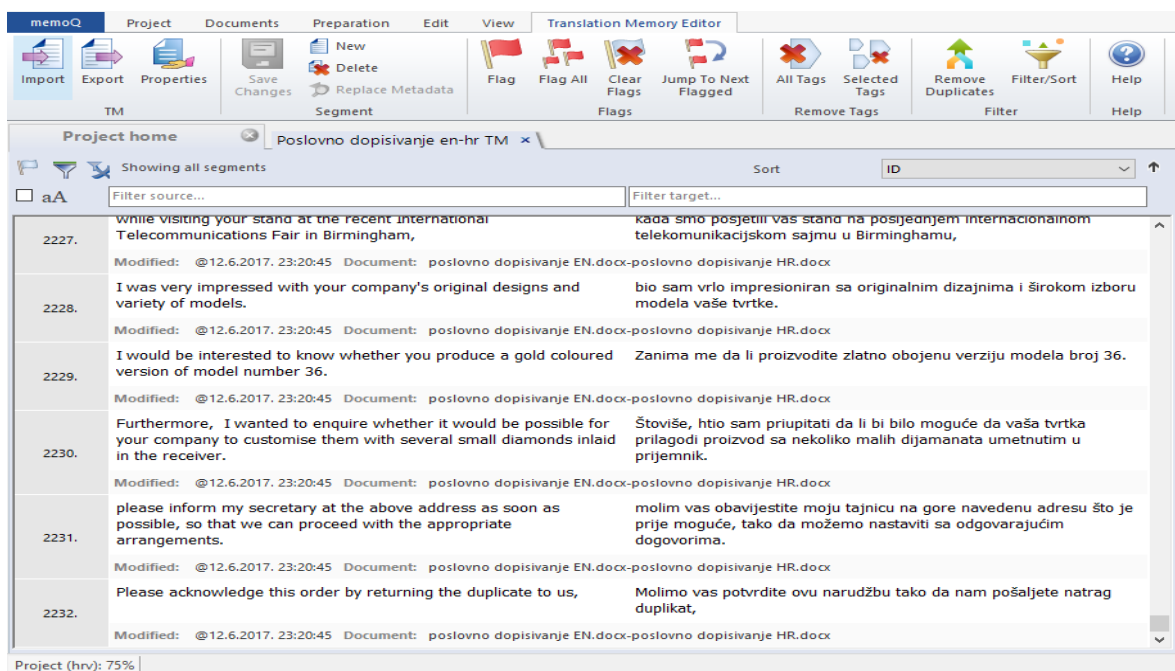
Oba su prikazana primjera rezultat pogrešne segmentacije na koju su utjecali interpunkcijski znakovi (ovdje je to bila dvotočka), praćeni s velikim početnim slovom, bilo u izvorniku ili prijevodu. Alat nailaženjem na točku, dvotočku, trotočku i ostale specifične interpunkcijske znakove, praćene velikim početnim slovom, segmente razdvaja na dva dijela, što ne mora biti slučaj i u prijevodu ako nakon određenog interpunkcijskog znaka ne slijedi veliko slovo i ne radi se o završetku rečenice.

Po završetku sravnjivanja, sve veze među segmentima moraju biti potvrđene (plava boja poveznice), a sljedeći je korak njihovo dodavanje u prijevodnu memoriju koja je kreirana pri postavljanju projekta opcijom *Export to TM* (Slika 3.).



Slika 3. Spremanje segmenata u prijevodnu memoriju

Rezultat sravnjivanja tekstova je prijevodna memorija *Poslovno dopisivanje en-hr TM.tmx*, s ukupno 2232 segmenta (Slika 4.).

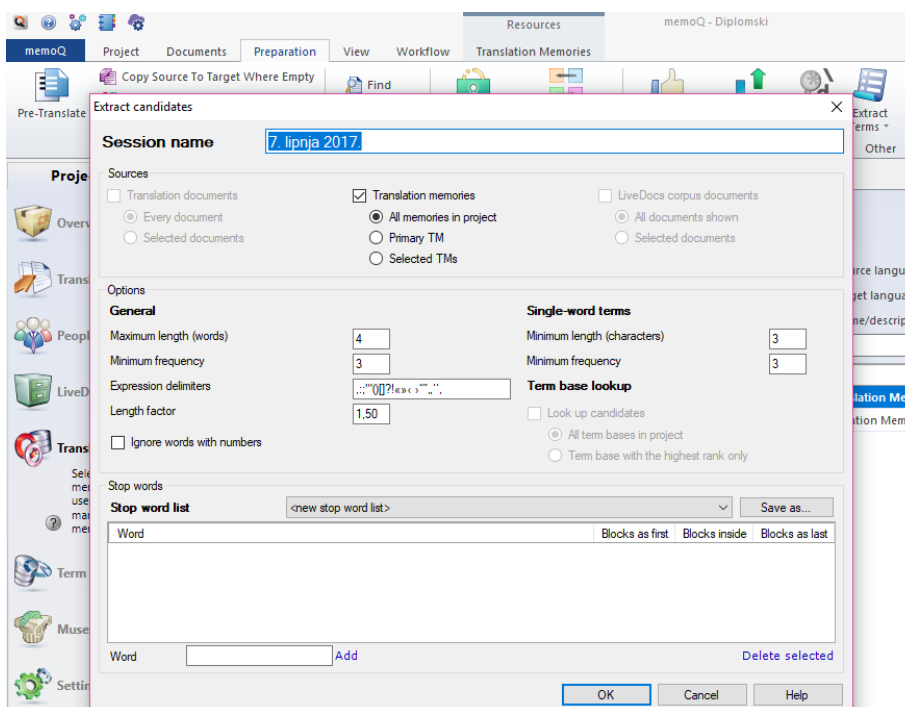


Slika 4. Prijevodna memorija *Poslovno dopisivanje en-hr TM*

5.3. Ekstrakcija terminologije

Drugi se dio praktičnog dijela ovog rada sastojao od automatske ekstrakcije terminologije koja će poslužiti kao resurs u analizi kvalitete prijevoda. Ekstrakcija terminologije se također radila u memoQ alatu, opcijom *Extract terms*. Prije samog početka ekstrakcije, u već kreirani projekt je potrebno dodati praznu terminološku bazu u koju će kasnije biti ubačeni termini. U ovome projektu terminološka baza je nazvana *Poslovno dopisivanje en-hr TB*. Ekstrakcija terminologije je napravljena nad prijevodnom memorijom dobivenom sravnjivanjem tekstova iz prethodnog zadatka. Za razliku od automatskog sravnjivanja, automatska ekstrakcija terminologije zahtijeva veću ljudsku intervenciju, s obzirom na to da se zasniva na faktoru čestote pojavljivanja riječi ili niza riječi izvornika i prijevodnog ekvivalenta.

Na Slici 5. su vidljive postavke ekstrakcije koje je moguće namjestiti prema potrebama određenog zadatka. Opcije koje je moguće namjestiti su maksimalna duljina termina izražena u riječima, minimalna frekvencija termina, popis znakova koji se ubrajaju među graničnike, faktor duljine, mogućnost ignoriranja termina s brojevima te na razini termina od jedne riječi, minimalna duljina riječi u znakovima i minimalna frekvencija pojavljivanja. U ovom zadatku su korištene već zadane opcije, vidljive na sljedećoj slici.



Slika 5. Ekstrakcija terminologije u memoQ alatu

Rezultat ekstrakcije termina je popis mogućih kandidata između kojih terminolog (ili bilo koja druga osoba koja radi ekstrakciju) odlučuje radi li se o stvarnim ili pogrešnim terminima. Samim time je ovaj zadatak najviše subjektivan, no alati još uvijek ne pružaju toliko dobre rezultate da bi se smanjila velika potreba za ljudskom intervencijom, pogotovo u slučaju flektivnog jezika poput hrvatskog. Duljina popisa mogućih kandidata za termine i vrijeme potrebno za prihvaćanje ili odbacivanje kandidata ovisit će o prethodno spomenutim postavkama ekstrakcije. Povećanje minimalne frekvencije pojavljivanja riječi ili niza riječi kandidata dat će kraći popis, no i mogućnost izostanka nekog stvarnog termina unutar domene, zbog manje frekvencije. Na Slici 6. je prikazan rezultat automatske ekstrakcije termina korištenjem zadanih postavki ekstrakcije. S lijeve su strane predloženi kandidati iz izvornika, a u desnu se stranu upisuje prijevod određenog termina na temelju pojavljivanja unutar prijevoda. Prijevodi predloženog termina iz originala su vidljivi u donjem dijelu ekrana. Zbog svoje visoke frekvencije pojavljivanja u jeziku, nepromjenjive vrste riječi (veznici, prijedlozi, članovi itd.) često dovode do lošije kvalitete rezultata ovakvih alata, stoga je potreba za ljudskom intervencijom veća.

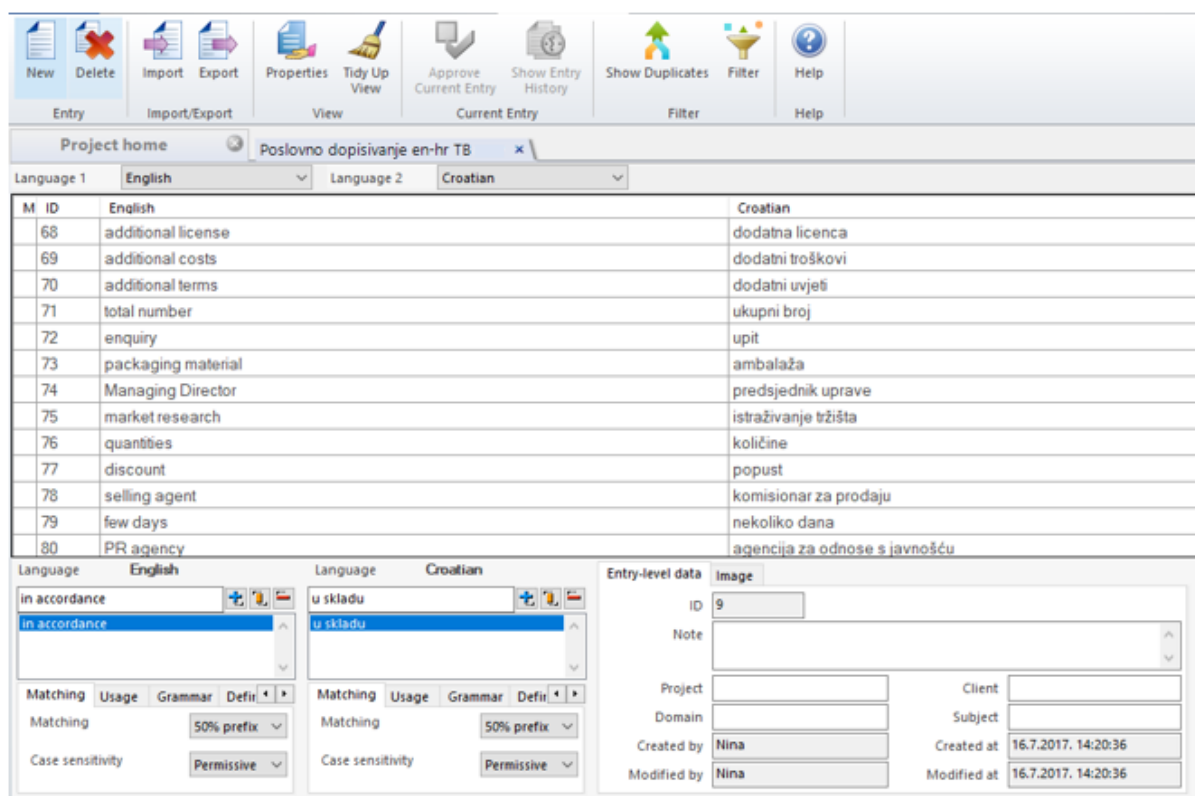
Score	#	H...	Status	Source	
2,89	45		Candidate	commission buyer	komisionar za kupnju
2,89	45		Candidate	as you	
2,89	45		Candidate	selling agent	komisionar za prodaju
2,89	45		Candidate	in this	
2,89	45		Candidate	copy of	
2,89	45		Candidate	a certain	
2,89	45		Candidate	details of	

Occurrences	Navigation: [Ctrl+Up]/[Ctrl+Down]	Go to [Ctrl+Space]	Term base results
svoj posao.			selling agent
It would certainly be worth your while to act as our selling agent in X.	Zasigurno bi Vam se isplatilo da preuzmete ulogu našeg komisionara za prodaju u X.		
I will be pleased to act as your selling agent here in X.	Rado ću preuzeti ulogu Vašeg komisionara za prodaju u X.		

Source: Poslovno dopisivanje en-hr TM.tmx (TranslationDoc)

Slika 6. Rezultat pokretanja ekstrakcije termina u memoQ-u

Ekstrakcija termina iz prijevodne memorije *Poslovno dopisivanje en-hr TM* dala je 5649 kandidata za termine između kojih je odabrano 85 termina koji su zatim dodani u terminološku bazu *Poslovno dopisivanje en-hr TB* (Slika 7.).



Slika 7. Terminološka baza Poslovno dopisivanje en-hr TB

Odabrani termini poslužit će kao digitalni resurs u analizi kvalitete prijevoda tekstova iz domene poslovnog dopisivanja. Korištenjem dobivene terminološke baze najčešćih pojavljivanja termina iz domene provjerit će se njihova konzistentnost i terminološka odstupanja u prijevodu. Za potrebe provjere terminologije, terminološka baza je izvedena iz memoQ alata u obliku Excel tablice s dva stupca termina, na engleskom i hrvatskom jeziku. Ovisno o alatu u kojem će se raditi provjera kvalitete prijevoda, format datoteke će se po potrebi mijenjati.

5.4. Analiza kvalitete prijevoda

Za analizu kvalitete prijevoda tekstova s engleskog na hrvatski jezik iz domene poslovnog dopisivanja su odabrana već ranije spomenuta četiri alata: memoQ, Verifika, QA Distiller i Xbench. Analiza kvalitete prijevoda sastojat će se od pregledavanja alatom pronađenih pogrešaka te njihova prebrojavanja. Analiza će pokazati koliko je alatima pronađenih stvarnih pogrešaka te koliko se kvaliteta prijevoda poboljšava njihovim ispravljanjem. U prvom dijelu analize bit će izračunati stupnjevi pogreške (engl. *level of mistakes*) svakog alata, a zatim će na temelju nađenih pogrešaka biti izračunata ocjena kvalitete (engl. *translation quality*).

5.4.1. Tipovi grešaka

Nazivi i grupe pogrešaka razlikuju se od alata do alata, no za potrebe ovog zadatka postavke provjere kvalitete su maksimalno usklađene kako bi se rezultati mogli što bolje usporediti, a neke vrste pogrešaka su izbačene iz analize jer su bile zanemarive i nisu bile dio postojećih provjera u svim alatima. U konačnici, odabrani tipovi pogrešaka za analizu kvalitete prijevoda su:³⁴

1. Pravopis – pravopisne pogreške tipa: ... *s obzirom na vioku kvalitetu robe ...*
2. Terminologija – pogrešna upotreba termina iz terminološke baze dobivene ekstrakcijom termina u jednom od prethodnih poglavlja, npr. *buying agent – komisionar za kupnju; Could I also act as a buying agent on your behalf?/ Bih li mogao za Vas raditi i kao komisionar za prodaju?*
3. Konzistentnost prijevoda – provjera dosljednosti u prijevodu (pronađen isti segment u izvorniku preveden na više različitih načina), npr. *All claims and entitlements on the part of the agent/representative are discharged on payment of the commission. – Plaćanjem provizije ispunjeni su svi zahtjevi zastupnika. / Plaćanjem provizije ispunjeni su svi troškovi.*
4. Konzistentnost izvornika – provjera dosljednosti u izvorniku (pronađeni različiti segmenti u izvorniku, a isti u prijevodu), nor. *Would you be willing to take on this responsibility? / Would you be prepared to take on this responsibility? – Biste li bili voljni preuzeti taj zadatak?*

³⁴ Uz svaki je tip pogreške naveden primjer pronađene greške tog tipa.

5. Početno slovo – provjera dosljednosti pisanja velikog ili malog početnog slova u segmentima izvornika i prijevoda, npr. *Yours faithfully. – srdačno.*
6. Dvostruki razmaci – greške dvostrukih razmaka među riječima, npr. *U svijetu ste poznati kao tvrtka _specijalizirana za pakiranje.*
7. Brojevi – pogreške vezane uz nedosljednosti među brojevima iz izvornika i iz prijevoda, kao i greške formatiranja brojeva koje ne odgovaraju pravilima pisanja brojeva u ciljnom jeziku, npr. *Otherwise, please forward us the amount owed in full by March 1st, 20--.* – *U suprotnom, molimo da uplatite ukupan iznos koji dugujete do 1.ožujka 2005.*
8. Pogreške interpunkcije – nedosljednost interpunkcijskih znakova s izvornikom ili nepravilna upotreba interpunkcije za ciljni jezik, npr. *To cover his expenses the agent/representative will recieve a monthly allowance of ...* - *Kako bi pokrio troškove, zastupnik će dobivati mjesečni paušal od....*
9. Navodnici – pogrešna upotreba navodnika za ciljni jezik (u hrvatskom jeziku pravilna upotreba navodnika je „ i ”), npr. *Mjesto rada može biti "jedan" resurs ili "grupa resursa" gdje je jedan raspored kreiran za redak.*
10. Ponovljene riječi – ponovljene iste riječi ili fraze u ciljnom jeziku, npr. *Ovaj ugovor o zastupanju najprije se najprije se zaključuje na probni rok od šest mjeseci.*
11. Razmaci oko znakova – pogrešna upotreba razmaka ispred ili iza interpunkcijskih ili drugih znakova (postotak, stupanj i sl.), npr. *(...) oslanja se na to da će vlasnici sezonskih ulaznica popuniti 25% sjedećih mjesta na svakom događaju...*
12. Pogreške URL-a – pogrešno preveden URL ili nedosljedan URL, kao i izostanak URL-a u ciljnom jeziku, npr. *Country availability, supported languages, supported operating systems and databases are provided through the Product Availability Matrix (PAM) published at www.servoce.sap.com/pam.* – *Raspoloživost države, podržani jezici, podržani operativni sustavi i baze podataka osigurani kroz Product Availability Matrix (PAM) objavljeni na www.service.sap.com/pam.*³⁵

³⁵ U primjeru je vidljivo da se radi o grešci u izvorniku (*servoce*), no ovdje je dano samo kao primjer pronađene greške nedosljednosti između URL-a izvornika i prijevoda i neće se računati kao greška u prijevodu.

5.4.2. Stupanj pogreške (engl. *Level of mistake*)

Za usporedbu uspješnosti alata računat će se stupanj pogreške za svaki alat. Formula prema kojoj će biti izračunat stupanj pogreške je sljedeća³⁶:

$$L = \frac{e \times 100}{d + n} ;$$

gdje su:

$$e = f + n;$$

f – alatom nađene pogrešno javljene greške;

n – alatom nepronađene stvarne greške;

d – sve alatom pronađene greške (stvarne i one pogrešne).

Analiza stvarnih i zanemarivih grešaka te računanje stupnja pogreške će pokazati korištenjem kojeg alata bi se kvaliteta prijevoda najviše pospješila. Taj će alat imati najmanji stupanj pogreške, tj. odnos stvarnih te alatom pogrešno javljenih grešaka.

U nastavku slijedi opis alata korištenih u analizi, rezultati svakog od alata te usporedba rezultata.

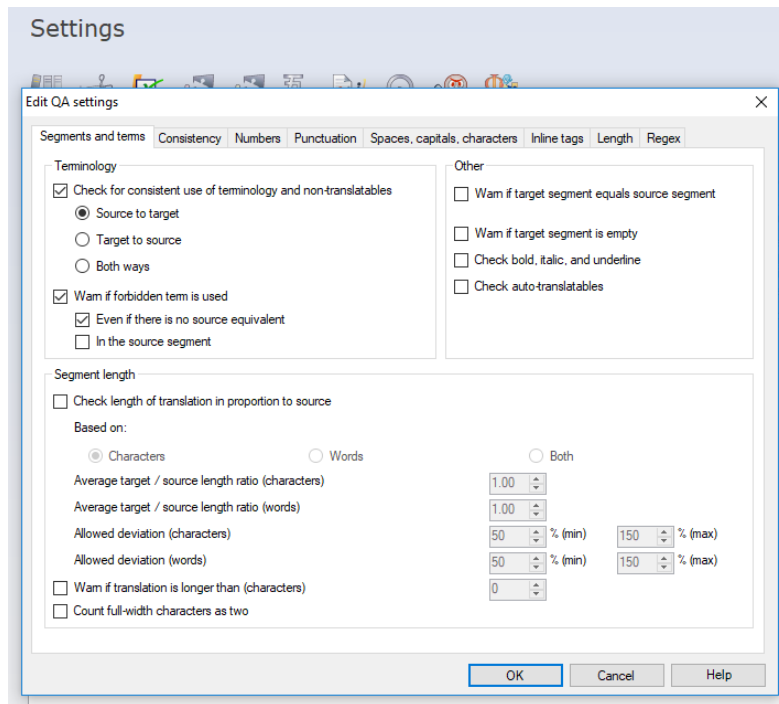
5.4.3. Alati

5.4.3.1. *memoQ*

memoQ je prvi alat koji je korišten za provjeru kvalitete prijevoda. Osiguranje kvalitete prijevoda (QA) je dio memoQ alata, stoga ga je jedino potrebno pokrenuti pritiskom na *Preparation – Quality Assurance*. Rezultat pokretanja QA-a je popis pronađenih grešaka između kojih kontrolor kvalitete odlučuje koje su stvarne pogreške, a koje se mogu ignorirati. Standardne provjere koje su dio kontrole kvalitete u alatu memoQ su provjera konzistentnosti, brojeva, interpunkcije, ali i pravopisa te konzistentnosti s terminološkom bazom. Korištenjem unaprijed zadanih provjera alat nije toliko precizan kao neki drugi (samostalni) QA alati, no njegova se uspješnost povisuje ukoliko se namjeste i dodatne, jezično specifične postavke, kao što su definiranje pravila za razmake oko interpunkcijskih

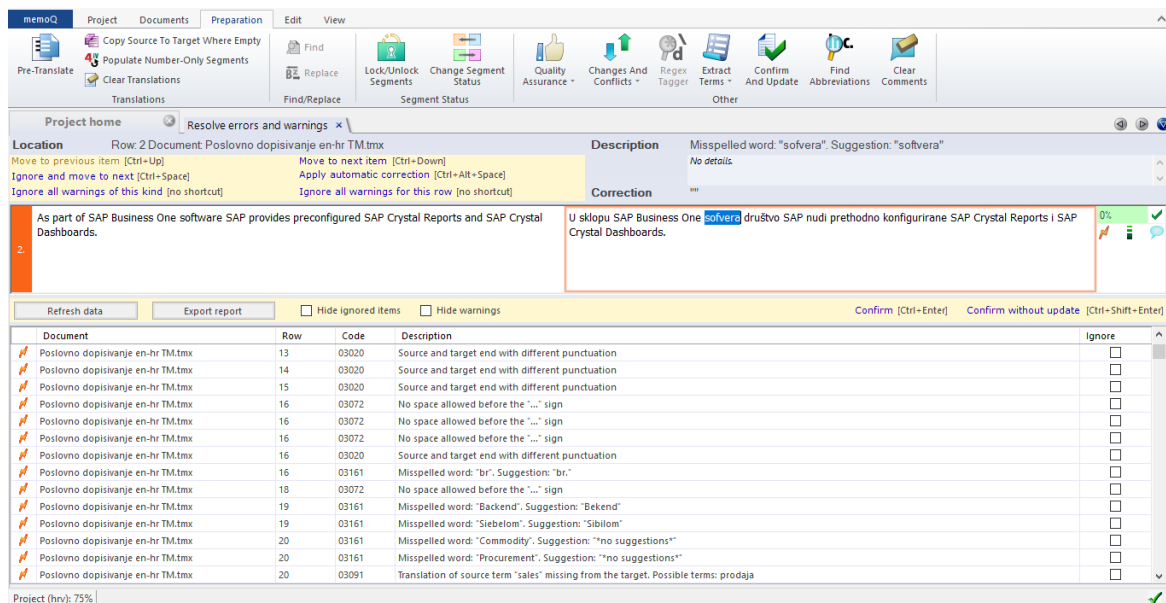
³⁶ Gural, Svetlana K.; Chemezov, Yan R. Analysis of Efficiency of Translation Quality Assurance Tools / Procedia - Social and Behavioral Sciences 154 (2014) 360 – 363 URL: <http://www.sciencedirect.com/science/article/pii/S1877042814056134> [10.9.2017.]

znakova, pisanje decimalnih brojeva te uključivanje provjere pravopisa za ciljni jezik, što također nije dio zadanih postavki. Za potrebe ovog istraživanja postavke su namještene onoliko koliko je to bilo potrebno za testiranje alata na osnovu ranije spomenutih 12 tipova grešaka.



Slika 8. Namještanje QA postavki - memoQ

Na Slici 9. je prikazan rezultat pokretanja QA-a.



Slika 9. QA u memoQ-u

U donjem dijelu prozora su izlistane greške sa svojim opisima, dok je u gornjem dijelu ekrana prikazan segment u kojem se pojavljuje određena greška, s izvornikom s lijeve strane i prijevodom s desne strane. Grešku je odmah moguće ispraviti u prozoru prijevoda ili označiti „*ignore*” u kvadratiću s desne strane. Alat ima opcije ignoriranja svih grešaka istog tipa, kao i svih grešaka u određenom retku prijevoda te automatskog ispravka. Po potrebi je moguće osvježiti rezultat, npr. kod unošenja izmjena u prijevod, te napraviti izvještaj i sakriti ignorirane greške ili upozorenja. Izvještaj QA rezultata sprema se u obliku *.html datoteke, s izlistanim greškama (ignorirane imaju precrtan broj segmenta u kojem se nalaze) i brojem grešaka određenog tipa.

5.4.3.1.1. Rezultati memoQ

Zbog različitosti u prikazima rezultata u izvještaju i bolje preglednosti, rezultati pojedinih alata s popisom promatranih grešaka će biti prikazani u tablicama, kao što je to Tablica 1. za alat memoQ. U tablici su, prema tipu pogreške, dani brojevi stvarnih grešaka koje je alat pronašao (engl. *True*, oznaka T), pogrešno nađenih grešaka (engl. *False*, oznaka F) i zbroj stvarnih i pogrešno nađenih grešaka (Σ).

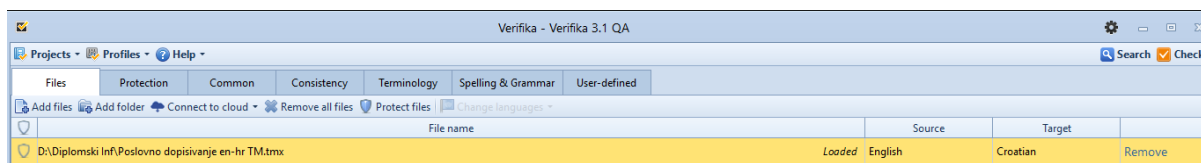
TIP POGREŠKE	memoQ 15		
	T	F	Σ
Pravopis	52	836	888
Terminologija	42	405	447
Konzistentnost prijevoda	24	19	43
Konzistentnost izvornika	10	15	25
Veliko slovo	1	19	20
Dvostruki razmaci	29	0	29
Brojevi	9	33	42
Interpunkcijske pogreške	109	82	191
Navodnici	0	0	0
Ponovljene riječi	1	0	1
Razmaci oko znakova	192	11	203
Pogreške URL-a	0	0	0
UKUPNO	469	1420	1889

Tablica 1. Rezultati provjere kvalitete u memoQ alatu

5.4.3.2. Verifika 3.1

Za razliku od alata memoQ, koji je prvenstveno CAT alat, Verifika je alat napravljen isključivo za provjeru kvalitete prijevoda. Mogućnosti Verifike rastu sa svakim novim ažuriranjem softvera, a verzija 3.1 koja je korištena u ovom radu ujedno je i najnovija.

Preuzimanje samog alata je besplatno, kao i prva dva tjedna korištenja, no duže korištenje zahtijeva unošenje licencnog broja, čija cijena ovisi o broju računala koja ju koriste i broju jezika provjere. Godišnja pretplata za korištenje na dvama računalima s dva jezika provjere iznosi 89 \$, dok godišnja pretplata za korištenje alata na više računala sa svim jezicima provjere iznosi 369 \$.³⁷ Provjera kvalitete prijevoda u alatu Verifika započinje dodavanjem datoteke spremne za provjeru. Alat podržava više oblika datoteke za provjeru kvalitete: tekstne datoteke (*.txt), SDL Trados dvojezične datoteke (*.sdlxliff), memoQ datoteke (*.xlf, *.mqxliff, *.mqxlz), MemSource dvojezične datoteke, Excel datoteke (*.xls, *.xlsx).³⁸



Slika 10. Kreiranje projekta (Verifika)

Kreiranje projekta u alatu počinje dodavanjem datoteke za QA (Slika 10.). Nakon dodavanja datoteke alat prepoznaje izvorni i ciljni jezik prema jezicima koji su namješteni u CAT alatu iz kojeg dolazi dvojezična datoteka te dodaje automatske provjere za taj jezik, u ovom slučaju hrvatski. Zadane je postavke i u ovom alatu moguće namjestiti u postavkama prije samog pokretanja provjere kvalitete.

Alat nudi sljedeće mogućnosti provjere:

- *Protection* – zaštita određenih segmenata, odnosno njihovog isključivanja iz procesa QA-a, uz mogućnost provjere dosljednosti s tim segmentima (primjerice, zaključanih segmenata, 100 % usklađenih segmenata s memorijom (engl. *100% match*), segmenata koji se slažu 100 % s memorijom i kontekstom iz memorije (engl. *Context match* ili *101% match*) i sl.); popis vrste segmenata koje je moguće isključiti iz provjere ovisi o alatu iz kojeg datoteka dolazi, npr. u slučaju memoQ dvojezične datoteke moguće je zaštititi *100% match*, *101% match*, zaključane, potvrđene i pregledane (engl. *Proofread*) segmente;³⁹
- *Common* – namještanje općih provjera (brojevi, razmaci, interpunkcija itd.);

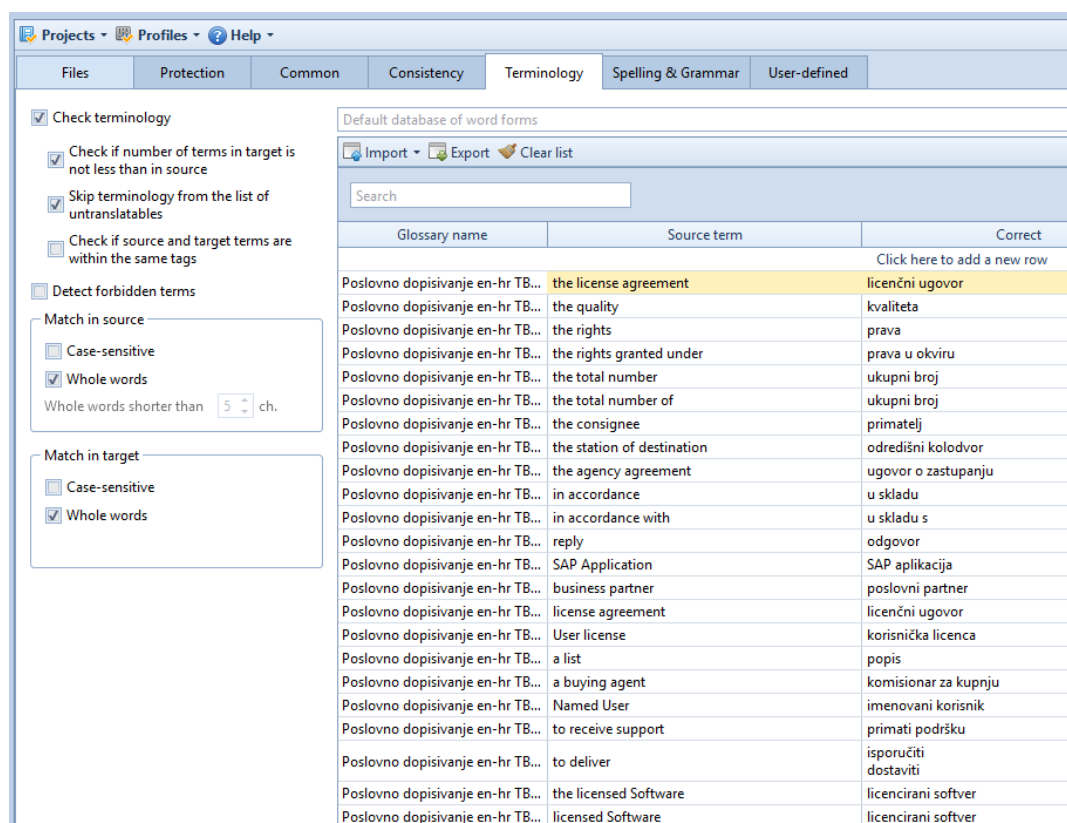
³⁷ Verifika pricing. Web page. URL: <https://e-verifika.com/pricing/> [2.8.2017.]

³⁸ Verifika: File Format Support. Web page. URL: http://help.e-verifika.com/general_information/file_format_support/ [2.8.2017.]

³⁹ Verifika: Protection. Web page. URL: http://help.e-verifika.com/detailed_information/protection/ [2.8.2017.]

- *Consistency* – provjera dosljednosti u izvorniku i ciljnom jeziku;
- *Terminology* – provjera terminologije;
- *Spelling & Grammar* – provjera pravopisa i gramatike (preko MS Worda ili Hunspella; ne funkcionira za sve jezike);
- *User Defined* – dodatne provjere koje je moguće ručno dodati u projekt.

Za potrebe QA-a u ovom su radu korištene zadane provjere za hrvatski jezik, uz uključivanje nekoliko dodatnih provjera, kako bi tipovi promatranih grešaka među alatima bili što sličniji za analizu. Primjerice, provjera termina nije dio zadanih provjera, stoga je u projektu bilo potrebno uključiti terminološku provjeru te dodati terminološku bazu u obliku Excel tablice (Slika 11.). Uvjet za ispravno prepoznavanje termina u alatu su popis termina u dva uzastopna stupca (prvi za termine u izvornom jeziku, a drugi za termine u ciljnom jeziku) te datoteka spremljena u nekom od alatom podržanih formata: *.xls, *.txt, *.csv, *.tbx, *.xml, *.sdltdb, *.xlsm.⁴⁰

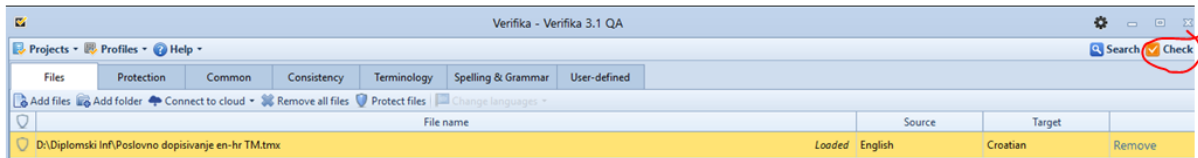


Slika 11. Dodavanje terminološke baze u projekt

⁴⁰ Verifika File Format Support. Web page. URL: http://help.e-verifika.com/general_information/file_format_support/ [2.8.2017.]

Nakon dodavanja terminološke baze u provjeru, projekt s namještenim postavkama i dodanim datotekama je moguće spremirati i ponovno pokrenuti u bilo kojem trenutku u slučaju prekida rada na projektu.

Provjera kvalitete prijevoda počinje klikom na „Provjeri” (engl. *Check*) u gornjem desnom kutu.



Slika 12. Pokretanje projekta (Verifika)

Rezultat pokretanja provjere kvalitete su popisi grešaka razvrstani prema ranije spomenutim grupama grešaka (*Common, Consistency, Terminology, Spelling & Grammar*). Za potrebe ovoga rada nisu dodane posebne korisničke provjere u dijelu *User Defined*, a provjera gramatike je također onemogućena, zbog nedostatka provjernika gramatike za hrvatski jezik u MS Wordu.

Greške se mogu pregledavati željenim redoslijedom, a ignoriraju se desnim klikom miša ili razmaknicom. Više grešaka ignorira se pritiskom na prvi željeni redak, držeći *Shift* te pritiskom na zadnji željeni redak desnim klikom miša. Sve greške u početku imaju crvenu početnu oznaku koja ignoriranjem postaje siva. Zbog lakšeg pregleda, prilikom provjere moguće je sakriti ignorirane ili neignorirane greške.

Common	Consistency	Terminology	Spelling	User-defined	Search	Files	Pending changes	Source segment	Target segment	File name	Position
> No space after punctuation (9)											
								bpt SAP Application Visualization User ept is a Named User who is solely aut	bpt Korisnik tipa SAP Application Visualization User ept imenovani je ko	Poslovno dopisivanje...	1752
								bpt SAP Application Standalone Visualization User ept is a Named User who	bpt Korisnik tipa SAP Application Standalone Visualization User ept ime	Poslovno dopisivanje...	1754
								James Blond joined the A.N.Y. Company in July 1998.	James Blond pridružio se A.N.Y. Company u srpnju 1998. godine.	Poslovno dopisivanje...	1851
								Otherwise, please forward us the amount owed in full by March 1st. 20--.	U suprotnom, molimo da uplatite ukupan iznos koji dugujete do 1. ožujka 20	Poslovno dopisivanje...	1873
> No space after sign (3)											
								We do, however, enclose a list of our member firms in the bbb business.	Međutim, u prilogu dostavljamo popis tvrtki koje su članice gospodarske gra	Poslovno dopisivanje...	522
								As our field worker for the . . . area you could earn up to ... per month.	Kao vanjski suradnik za područje, mjesečno možete zaraditi i do...	Poslovno dopisivanje...	856
								Owing to an unfortunate delay, I regret to inform you that the delivery can only	Uslijed stanovitog nezgodnog zakašnjenja, moram Vas, nažalost, obavijestiti	Poslovno dopisivanje...	1361
> No space before sign (10)											
<input checked="" type="checkbox"/> [Ignored] <> ""											
Source segment We do, however, enclose a list of our member firms in the bbb business.											
Target segment Međutim, u prilogu dostavljamo popis tvrtki koje su članice gospodarske grane bbb.											
Suggested autofix Apply (ALT+A) Međutim, u prilogu dostavljamo popis tvrtki koje su članice gospodarske grane bbb.											

Slika 13. Rezultat provjere kvalitete (Verifika)

Kao što je vidljivo na Slici 13., unutar svake grupe grešaka postoje i podgrupe vezane uz različite tipove provjera (npr. nedostatak razmaka prije / nakon interpunkcijskog znaka (engl. *No space after punctuation*) u općim (engl. *Common*) greškama itd.). Izvještaj rezultata provjere kvalitete prijevoda (samo određene grupe grešaka ili svih grešaka) može se spremirati u *.xls formatu datoteke, a prilikom spremanja rezultata moguće je označiti želimo li u izvještaj uključiti i ignorirane greške. U tom slučaju će one biti prikazane u donjem dijelu Excel tablice pod naslovom Ignorirane greške (engl. *Ignored errors*), dok će stvarne greške biti prikazane na početku tablice. Verifika omogućava ispravljanje grešaka u segmentima na vrlo jednostavan način. Osim što za većinu tipova grešaka nudi automatski ispravak (engl. *Suggested autofix*) na temelju spremljenih postavki za ciljani jezik, u segment prijevoda (engl. *Target segment*) je moguće unijeti bilo koji željeni ispravak. Širi kontekst segmenta može se vidjeti duplim klikom na njega, a greške se mogu ispravljati i u tom prikazu. Napravljeni ispravci se spremaju u kartici izmjena na čekanju (engl. *Pending changes*). Jedna od korisnih mogućnosti alata je spremanje popisa izmjena na čekanju u obliku Excel datoteke prije spremanja samih izmjena u dvojezičnoj datoteci. Taj popis može poslužiti voditelju projekta u agenciji ili klijentu kao uvid u unesene izmjene u prijevodu.

5.4.3.2.1. Rezultati – Verifika

Konačni rezultat pronađenih grešaka u alatu Verifika dan je u Tablici 2.

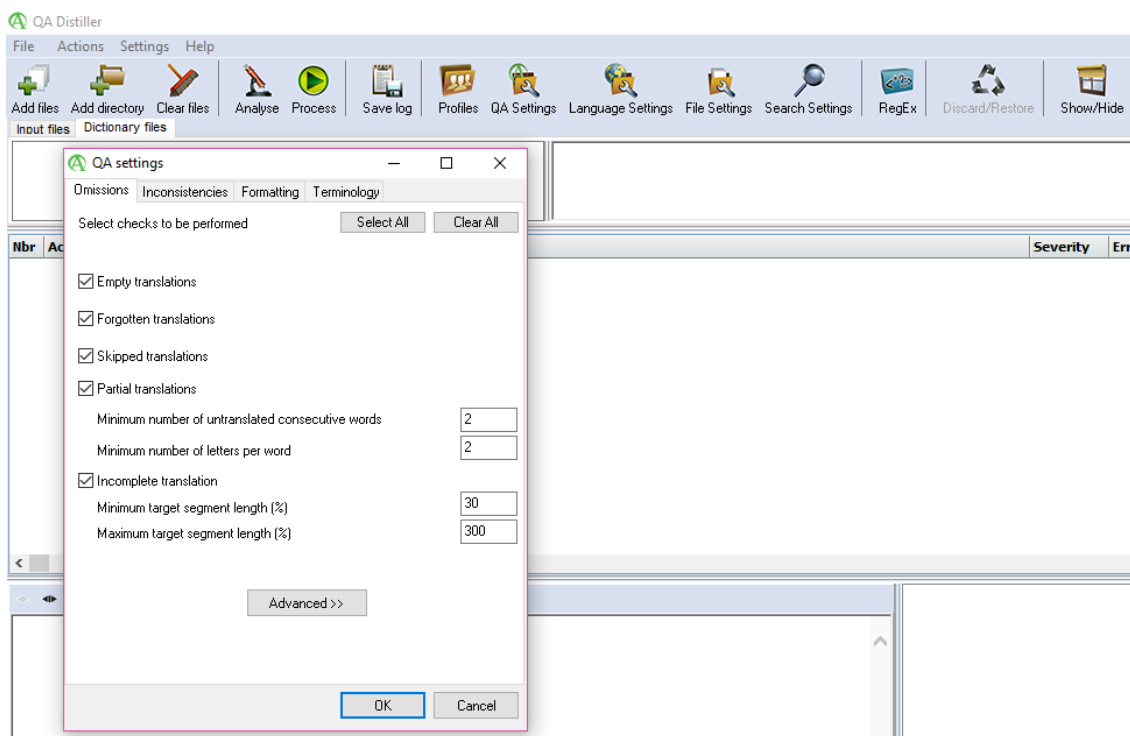
TIP POGREŠKE	Verifika 3.1		
	T	F	Σ
Pravopis	53	157	210
Terminologija	42	212	254
Konzistentnost prijevoda	24	19	43
Konzistentnost izvornika	10	15	25
Veliko slovo	1	20	21
Dvostruki razmaci	29	0	29
Brojevi	9	140	149
Interpunkcijske pogreške	96	62	158
Navodnici	47	0	47
Ponovljene riječi	2	0	2
Razmaci oko znakova	204	14	218
Pogreške URL-a	0	1	1
UKUPNO	517	640	1157

Tablica 2. Rezultati provjere kvalitete u alatu Verifika

5.4.3.3. QA Distiller 9

QA Distiller (skraćeno: QAD) je, poput Verifike, samostalni alat za provjeru kvalitete prijevoda. Najnovija je verzija br. 9, ujedno korištena u ovome radu. Vrijednost licence je 249, 1000 ili 2500 €, ovisno o broju korisnika i jezika koje podržava (od 3 do 90 ili više jezika).⁴¹ Besplatna verzija traje sveukupno tri tjedna i sadrži sve funkcionalnosti kao i plaćena verzija. Kontrola kvalitete počinje dodavanjem dvojezične datoteke u jednom od alatom podržanih formata.⁴² Nakon dodavanja dvojezične datoteke s izvornim tekstom i prijevodom, u provjeru se može uključiti i terminološka baza, no u slučaju QA Distillera ona mora biti obrađena na specifičan način. Sam alat podržava *.tbx i *.xml datoteke, no u slučaju npr. Excel datoteke, ona mora biti pretvorena u *.dic datoteku koju alat podržava. Konverzija datoteke odvija se u samom alatu. Nakon što je spremljena terminološka baza u *.dic formatu, ona može biti uvrštena u svaki QA, bez ponavljanja procesa konverzije.

Grupe provjera koje QA Distiller nudi su greške propusta (engl. *Omissions*), nedosljednosti (engl. *Inconsistencies*), formatiranja (engl. *Formatting*) i terminologije (engl. *Terminology*). Kao i u ostalim alatima, postavke je moguće ručno namjestiti, a za potrebe praktičnog dijela u ovom su radu korištene zadane postavke, uz određene izmjene, kako bi rezultati analize bili što sličniji navedenoj vrsti provjere.



Slika 14. Postavke provjere kvalitete u QAD-u

⁴¹ QA Distiller. Buy. Web page. URL: <http://www.qa-distiller.com/en/buy> [2.8.2017.]

⁴² QA Distiller. Supported file formats. Web page. URL: <http://www.qa-distiller.com/en> [2.8.2017.]

Slika 15. prikazuje rezultat pokretanja QA-a u alatu. U sredini ekrana su izlistane pronađene greške, u donjem lijevom kutu su prikazani izvorni i ciljni segment, a u donjem desnom kutu njihova pozicija u dokumentu.

Idr	Act	Source	Type	Translation	Sev	Error	File
149		and such Use shall be subject to the	Terminology	„a to korištenje podliježe „Licenci imenovanog korisnika“ i „Licenci paketa	5	Dictionary term User license was found in the source, but there	D:\Diplomsk
122		hereof with respect to applicability of	Terminology	„a u vezi primjenjivosti licenci imenovanog korisnika	5	Dictionary term Named User was found in the source, but there	D:\Diplomsk
138		A) that (i) adds any new, independent	Terminology	koji nije SAP u svrhu kreiranja novog sustava zapisa za te podatke ili metapodatke	5	Dictionary term SAP Application was found in the source, but th	D:\Diplomsk
153		A) that (i) adds any new, independent	Pattern match	koji nije SAP u svrhu kreiranja novog sustava zapisa za te podatke ili metapodatke	5	Target sentence contains regular expression string 'iii'	D:\Diplomsk
147		If your trial shipment is to our satisfa	Terminology	„Ako budemo zadovoljni probnom isporukom, možete računati s većim narudžbama	5	Dictionary term trial shipment was found in the source, but ther	D:\Diplomsk
140		Unless you achieve the minimum sa	Terminology	minimalni promet, bit ćemo prisiljeni trenutno otkazati Vaš ugovor o zastupanju	5	Dictionary term terminate the agency agreement was found in t	D:\Diplomsk
133		Unless you achieve the minimum sa	Terminology	minimalni promet, bit ćemo prisiljeni trenutno otkazati Vaš ugovor o zastupanju	5	Dictionary term sales was found in the source, but there is no	D:\Diplomsk
150		Should you have any queries please	Pattern match	našem komunikacijskom centru na broj bbb ili našem internetskom servisu na wwwbbb	5	Target sentence contains regular expression string 'bbb'	D:\Diplomsk
618		If a runtime database is licensed	Punctuation	„Ako je baza podataka vremena izvođenja licencirana	2	Source and target have a different end of sentence punctuation	D:\Diplomsk
414		If a runtime database is licensed	Inconsistent tra	„Ako je baza podataka vremena izvođenja licencirana	5	Identical source segment has multiple translations	D:\Diplomsk
146		Insofar as the consignee has provide	Terminology	sklopi ugovor o prijevozu, navedenom otpremniku bit će povjerena isporuka pošiljke	5	Dictionary term the station of destination was found in the sou	D:\Diplomsk
61		If licensee requires more than the m	Partially forgot	kao korisnici tipa SAP Business One Professional ili Limited Professional User	2	Untranslated segment found	D:\Diplomsk
60		If licensee requires more than the m	Partially forgot	kao korisnici tipa SAP Business One Professional ili Limited Professional User	2	Untranslated segment found	D:\Diplomsk

Partially forgotten translation - Severity 2

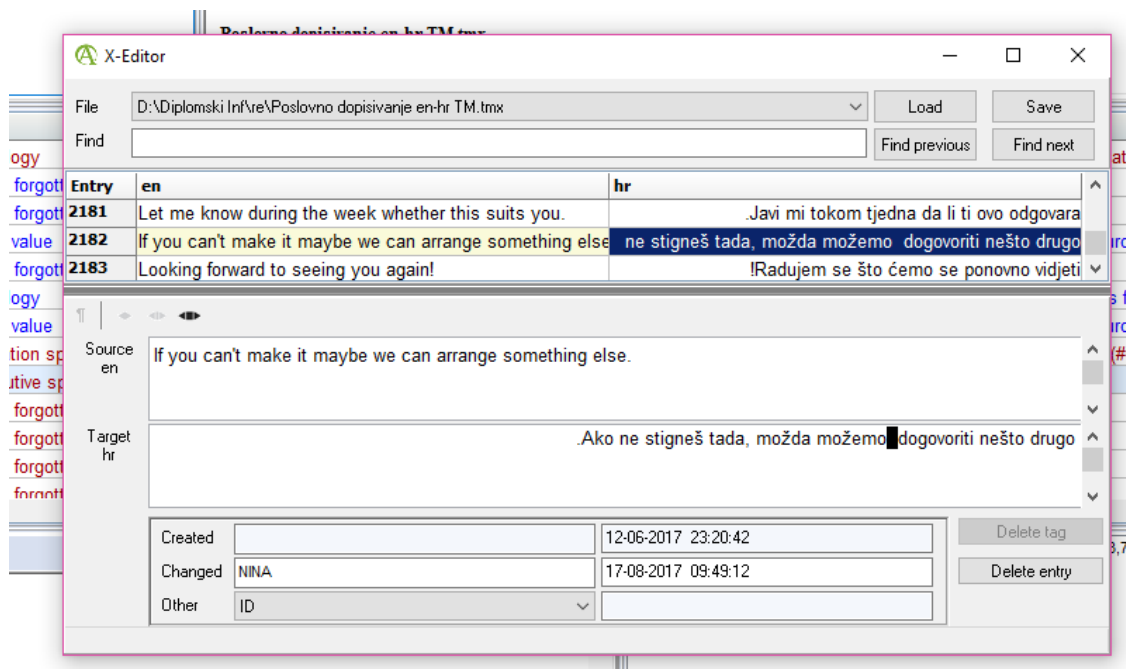
Untranslated segment found

Source If licensee requires more than the maximum number of five (5) SAP Business One Starter Package Users, all users need to be licensed as SAP Business One Professional or Limited Professional User.

Target Ako korisnik licence zahtijeva više od maksimalnog broja od pet (5) korisnika tipa SAP Business One Starter Package Users, svi korisnici trebaju biti licencirani kao korisnici tipa SAP Business One Professional ili Limited Professional User

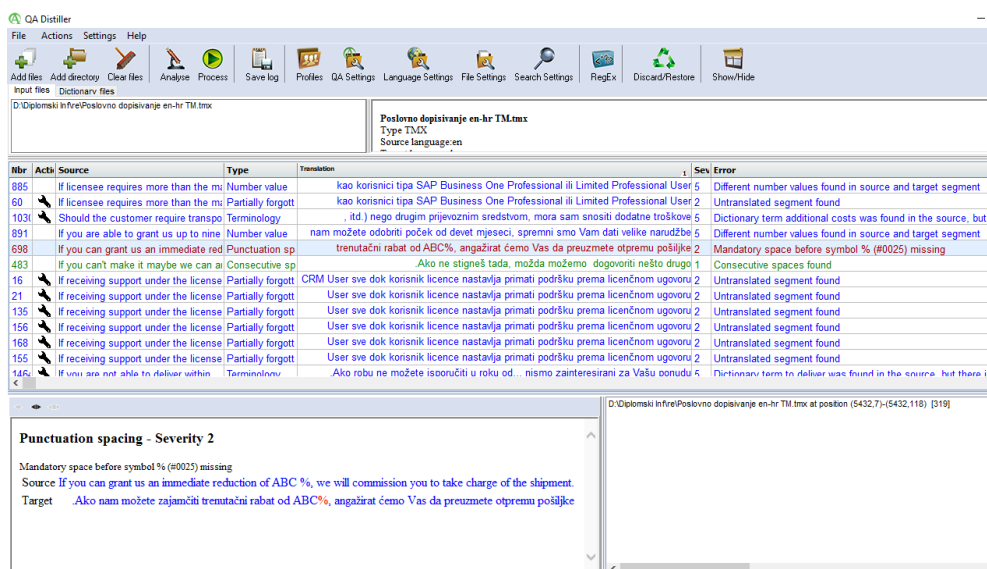
Slika 15. Pronađene greške u QAD-u

Ispravljanje grešaka moguće je dvostrukim klikom na segment s javljenom greškom te uređivanjem segmenta u X-Editoru, kao što je to prikazano na Slici 16. Klikom na Spremi (engl. Save) sprema se u datoteci unesena izmjena.



Slika 16. Ispravljanje grešaka u QAD-u

Prilikom njihova pregledavanja, greške se ignoriraju desnim klikom miša (plava boja segmenta), označavaju kao riješene (zelena boja segmenta) ili ostavljaju kao neriješene (crvena boja segmenta).



Slika 17. Pregledavanje pronađenih grešaka u QAD-u

Prikaz grešaka je moguće filtrirati prema različito označenim tipovima grešaka te spremiti izvještaj s pronađenim, ispravljenim i/ili ignoriranim greškama u *.txt, *.html, *.xml ili *.xls formatu.

5.4.3.3.1. Rezultati QA Distiller

Rezultati provjere kvalitete dani su u Tablici 3.

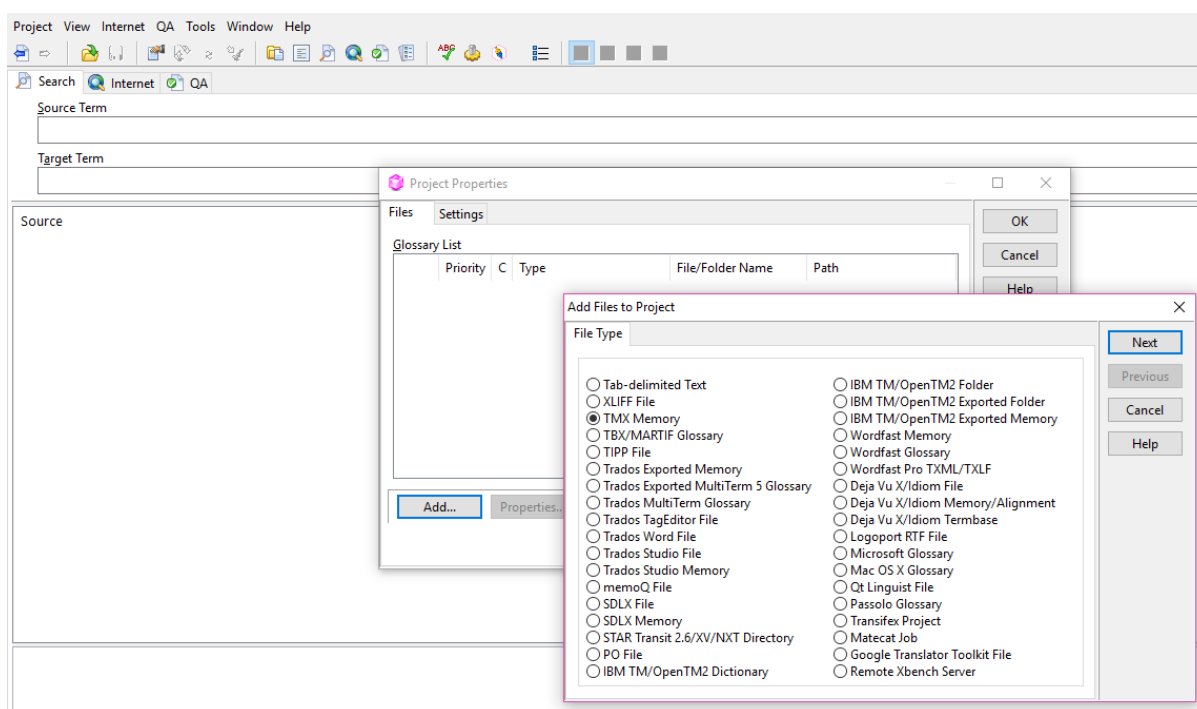
TIP POGREŠKE	QA Distiller 9.1.5		
	T	F	Σ
Pravopis	0	0	0
Terminologija	42	436	478
Konzistentnost prijevoda	24	19	43
Konzistentnost izvornika	9	15	24
Veliko slovo	1	19	20
Dvostruki razmaci	29	0	29
Brojevi	9	26	35
Interpunkcijske pogreške	96	57	153
Navodnici	47	0	47
Ponovljene riječi	0	0	0
Razmaci oko znakova	199	13	212
Pogreške URL-a	0	0	0
UKUPNO	456	585	1041

Tablica 3. Rezultati provjere kvalitete u alatu QA Distiller

5.4.3.4. Xbench 3.0

Xbench je još jedan samostalni alat za provjeru kvalitete prijevoda. Ono što ga razlikuje od ostalih alata korištenih u ovom radu je to što postoji verzija alata koja je besplatna⁴³, iako je u ovom istraživanju korištena verzija 3.0 (besplatna 30 dana korištenja alata, neovisno o kalendarski proteklim danima)⁴⁴. Xbench podržava velik broj formata dvojezičnih datoteka za provjeru, kao i terminoloških baza: *.txt, *.tsv, *.utx, *.xlf, *.xlif, *.xliff, *.xlz, *.mqxzlz, *.mqxliff, *.mxliff, *.tmx, *.xml, *.tbx, *.mtf te mnoge druge.⁴⁵ Osim terminološke provjere, alat omogućuje i provjeru pravopisa za 48 jezika, koristeći Hunspell i MS Word rječnike koji moraju biti instalirani na računalu.

Provjera kvalitete prijevoda počinje dodavanjem dvojezične datoteke u projekt, u ovom slučaju je to bila *.tmx dvojezična datoteka *Poslovno dopisivanje en-hr TM*.



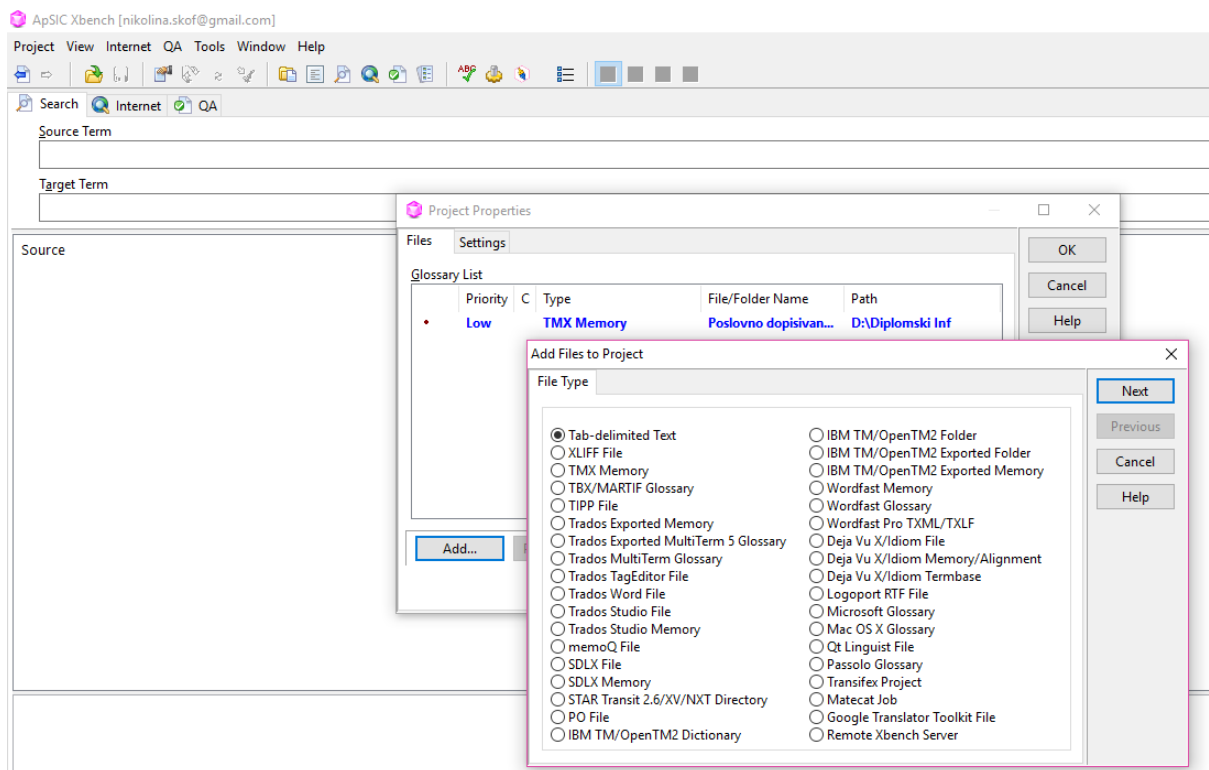
Slika 18. Dodavanje datoteke u projekt

Nakon dodavanja dvojezične datoteke za provjeru, dodaje se i terminološka baza u jednom od alatom podržanih formata. U ovom radu korištena je *.txt terminološka baza u kojoj su termini izvornog i ciljnog jezika odvojeni tabulatorom.

⁴³ Verzija 2.9, dostupna na: <https://www.xbench.net/index.php/download> [2.8.2017.]

⁴⁴ Cijena godišnje licence za program verzije 3.0 je €99 + VAT, URL: <https://www.xbench.net/index.php/store/order-xbench> [2.8.2017.]

⁴⁵ Popis alatom podržanih datoteka, dostupan na: <https://docs.xbench.net/user-guide/overview/> [2.8.2017.]

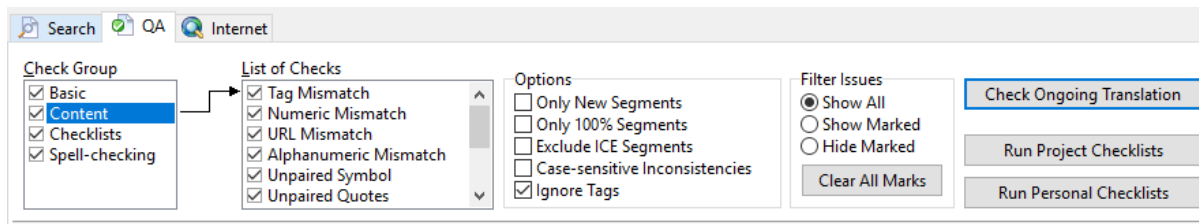


Slika 19. Dodavanje terminološke baze u projekt

Prilikom dodavanja datoteke za provjeru i terminološke baze moguće je definirati prioritet svake dodane datoteke (engl. *Priority*): nizak (engl. *Low*), srednji (engl. *Medium*), visok (engl. *High*) prioritet. Različito postavljene razine prioriteta imaju različite boje prikaza rezultata u opciji pretraživanja. Za ovo istraživanje postavljen je nizak prioritet za dvojezičnu *.tmx datoteku i visok prioritet za terminološku bazu. Osim namještanja prioriteta, potrebno je definirati vrstu datoteke koja će se provjeravati. Dvojezična *.tmx datoteka je ovdje postavljena kao prijevod (engl. *Ongoing translation*), a terminološka baza kao lista termina (engl. *Key Terms*).

Slika 20. prikazuje vrste provjera koje su dostupne u alatu Xbench. One su podijeljene u nekoliko skupina. U osnovne provjere spadaju provjere neprevedenih segmenata, nedosljednosti te jednakih segmenata u izvorniku i prijevodu. U drugu skupinu provjera spadaju provjera oznaka (engl. *tag*), brojeva, URL-a, neuparenih simbola i navodnika, višestruke praznine, ponovljenih riječi, nepronađenih riječi pisanih velikim slovima ili kombinacijom velikih i malih slova u ciljnom jeziku. U treću se skupinu ubrajaju tzv. Liste

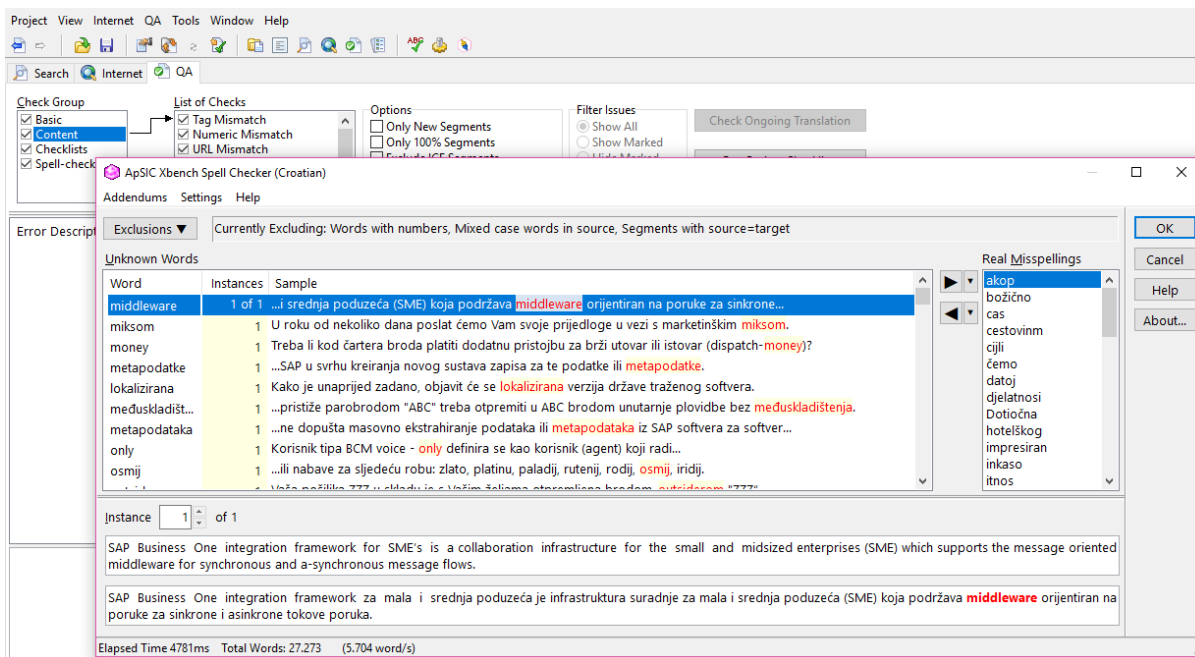
provjera (engl. *Checklists*), u koje se mogu spremi posebno namještene opcije pretrage.⁴⁶ Zadnju skupinu provjera čini provjera pravopisa.



Slika 20. Vrste provjera u alatu Xbench

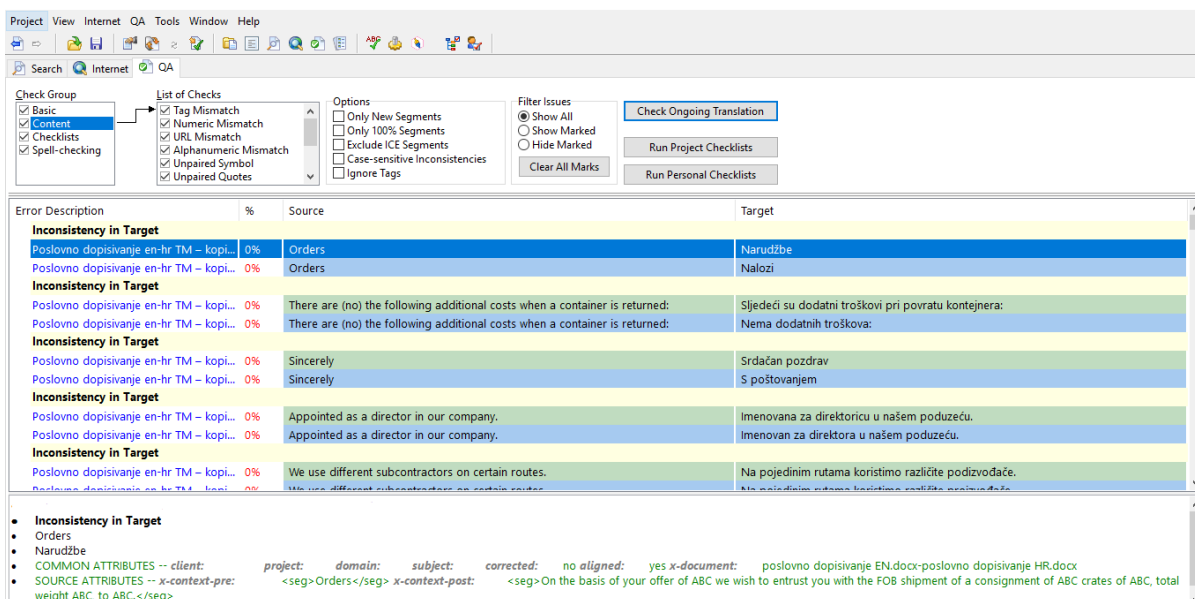
Osim uključivanja određenih vrsta provjere, moguće je odabrati želimo li provjeru samo novih segmenata (segmenata koji nisu povučeni iz memorije ili uređivani na temelju postojećeg sličnog prijevoda u memoriji), 100 % usklađenih segmenata s memorijom ili ICE segmenata (segmenti koji se slažu 100 % s memorijom i kontekstom iz memorije). Osim toga, omogućeno je i uključivanje provjere dosljednosti uzimajući u obzir velika i mala slova te opcija ignoriranja oznaka (engl. *Ignore tags*). Provjera kvalitete prijevoda sastojala se od prve grupe provjera (engl. *Basic*), druge skupine provjera (engl. *Content*), provjere pravopisa (engl. *Spell-checking*) te prethodno dodane terminološke provjere. Nakon pokretanja QA-a klikom na *Check Ongoing Translation*, prvo se pojavljuju pravopisne greške koje se potvrđuju (u slučaju da se stvarno radi o grešci) ili ne potvrđuju (u slučaju da se ne radi o grešci). Kad se potvrdi pogrešno napisana riječ, ona se svrsta u grupu stvarnih pravopisnih pogrešaka (engl. *Real Misspellings*), a klikom na OK te se greške pojavljuju na popisu zajedno s ostatkom QA izvještaja.

⁴⁶ Detaljnije o Listama provjere u poglavlju *Dodatne mogućnosti alata*.



Slika 21. Pronađene pravopisne greške

Greške u QA izvještaju su izlistane kao što je prikazano na Slici 22.



Slika 22. Rezultati provjere kvalitete u alatu Xbench

Greške se redom pregledavaju i označavaju (Ctrl + M) te dobivaju oznaku „X”, a moguće je i filtrirati prikaz prema označenim ili neoznačenim greškama. Za potrebu šireg konteksta, klikom na *Vidi kontekst* (engl. *See Context*), dobiva se pogled na okolinu segmenta s nađenom greškom. Sam alat nema mogućnost ispravljanja grešaka, već se klikom na *Uredi izvor* (engl. *Edit Source*) otvara datoteka u programu u kojem je nastala, s kursorom pozicioniranim u segmentu s javljenom greškom iz QA alata. Izvještaj rezultata provjere

kvalitete može se spremati u *.xls, *.txt, *.html i *.xml formatu. Nedostatak alata Xbench je nemogućnost ignoriranja greške unutar samo jednog segmenta u slučaju kad se ona pojavljuje u više segmenata. Primjerice, kratica *br.* ubrojena je u pravopisne greške kad se pojavljuje bez točke, no u Xbenchu nije moguće ignorirati grešku samo u segmentima u kojima je ona napisana s točkom, jer se sve pronađene pojave ubrajaju u jednu grešku. Ista stvar je i s nađenim terminološkim greškama. U nekim segmentima se nije radilo o grešci, jer je termin samo korišten u drugom padežu od onog u terminološkoj bazi, dok je u drugim segmentima stvarno korišten pogrešan termin. Iz tog je razloga bilo najteže prebrojati stvarne greške pronađene alatom Xbench, no konačni rezultati dani su u nastavku.

5.4.3.4.1. Rezultati Xbench

TIP POGREŠKE	Xbench 3.0		
	T	F	Σ
Pravopis	53	227	280
Terminologija	42	392	434
Konzistentnost prijevoda	23	19	42
Konzistentnost izvornika	10	15	25
Veliko slovo	0	0	0
Dvostruki razmaci	25	0	25
Brojevi	5	27	32
Interpunkcijske pogreške	0	0	0
Navodnici	1	0	1
Ponovljene riječi	1	0	1
Razmaci oko znakova	0	0	0
Pogreške URL-a	2	0	2
UKUPNO	162	680	842

Tablica 4. Rezultati provjere kvalitete u alatu Xbench

5.4.4. Analiza rezultata QA alata

U prethodnom poglavlju alati su detaljnije opisani s tehničke strane, zajedno uz mogućnosti koje nude u provjeri kvalitete prijevoda. Na kraju opisa postupka QA-a u svakom od alata prikazane su tablice s rezultatima pojedinih alata i pronađenih grešaka u procesu provjere kvalitete prijevoda. Za usporedbu alata prvo će biti izračunat stupanj pogreške (L) za svaki alat (Tablica 5.), a zatim će rezultati biti pobliže analizirani, uzimajući u obzir sve uočene prednosti i nedostatke proučavanih QA alata.

TIP POGREŠKE	t _{max}	memoq 15				Verfika 3.1				QA Distiller 9				Xbench 3.0			
		t	f	d	n	t	f	d	n	t	f	d	n	t	f	d	n
Pravopis	53	52	836	888	1	53	157	210	0	0	0	0	53	53	227	280	0
Terminologija	42	42	405	447	0	42	212	254	0	42	436	478	0	42	392	434	0
Konzistentnost prijevoda	24	24	19	43	0	24	19	43	0	24	19	43	0	23	19	42	1
Konzistentnost izvornika	10	10	15	25	0	10	15	25	0	9	15	24	1	10	15	25	0
Veliko slovo	1	1	19	20	0	1	20	21	0	1	19	20	0	0	0	0	1
Dvostruki razmaci	29	29	0	29	0	29	0	29	0	29	0	29	0	25	0	25	4
Brojevi	9	9	33	42	0	9	140	149	0	9	26	35	0	5	27	32	4
Interpunkcijske pogreške	109	109	82	191		96	62	158	13	96	57	153	13	0	0	0	109
Navodnici	47	0	0	0	47	47	0	47	0	47	0	47	0	1	0	1	46
Ponovljene riječi	2	1	0	1	1	2	0	2	0	0	0	0	2	1	0	1	1
Razmaci oko znakova	204	192	11	203	12	204	14	218	0	199	13	212	5	0	0	0	204
Pogreške URL-a	2	0	0	0	2	0	1	1	2	0	0	0	2	2	0	2	0
UKUPNO	532	469	1420	1889	63	517	640	1157	15	456	585	1041	76	162	680	842	370
STUPANJ POGREŠKE		75,97%				55,89%				59,18%				86,63%			

Tablica 5. Konačni rezultati QA alata

Tablica 5. prikazuje konačne rezultate alata. U prvom stupcu su nabrojene proučavane greške prema tipovima, a u drugom stvaran broj grešaka za svaki tip (t_{max}). U nastavku su dani brojevi grešaka koje su alati pronašli, i to:

t – broj stvarnih grešaka,

f – broj zanemarivih grešaka,

d – ukupan broj nađenih grešaka ($d = t + f$),

n – broj grešaka koje alat nije pronašao ($n = t_{max} - t$).

Stupanj pogreške L izračunat je prema formuli preuzetoj iz (Gural, Chemezov, 2014)⁴⁷ koja se koristi u studiji usporedbe nekoliko računalnih QA programa:

$$L = \frac{e \times 100}{d + n};$$

gdje je $e = f + n$.

Rezultati pokazuju sljedeće stupnjeve pogrešaka:

$$\text{memoQ} - L = 75,97 \%,$$

$$\text{Verifika} - L = 55,89 \%,$$

$$\text{QA Distiller} - L = 59,18 \%,$$

⁴⁷ Gural, Svetlana K.; Chemezov, Yan R. Analysis of Efficiency of Translation Quality Assurance Tools / Procedia - Social and Behavioral Sciences 154 (2014) 360 – 363 URL: <http://www.sciencedirect.com/science/article/pii/S1877042814056134> [10.9.2017.]

Xbench – L = 86,63 %.

Najboljim alatom za provjeru kvalitete prijevoda se pokazala Verifika sa stupnjem pogreške 55,89 %, a slijede ju QA Distiller, memoQ te na kraju Xbench.

U usporedbi s ostalim alatima, Verifika je jedina pronašla barem jednu pogrešku svakog tipa, unatoč tome što je u slučaju provjere URL-a pronašla jednu krivo javljenu grešku, a dvije stvarne nije. Razlog tome je što alat ima mogućnost provjere URL adrese uzimajući u obzir samo njezinu konzistentnost između izvornika i prijevoda (*www.servo~~o~~ce.sap.com/pam* – *www.service.sap.com/pam*, ali je greška ignorirana jer se ona u stvari nalazi u izvorniku), no ne i njezin izostanak u prijevodu (Xbench je jedini pronašao dva segmenta u kojima se URL adrese nalaze u segmentu izvornika, ali ih nema u prijevodu).

Osim spomenute provjere URL adrese, Verifika je ponudila najviše vrsta provjere i pronašla najveći broj stvarnih grešaka (517) te najmanji broj nepronađenih grešaka (15).

Nepronađene greške su bile dvije već spomenute URL adrese te one vezane uz interpunkcijske znakove (dupli interpunkcijski znakovi i nedosljednost interpunkcijskih znakova u izvorniku i prijevodu), u čijoj se provjeri najbolji pokazao alat memoQ.

Unatoč tome što i Verifika i QA Distiller nude različite mogućnosti provjere interpunkcije, nisu pronašli npr. duple interpunkcijske znakove kad se radi o različitim znakovima (*Prihvaćat će se službeno priznati troškovi u iznosu do....* – znak na kraju je trotočka i još jedna nepotrebna točka), a QAD nije pronašao niti nizove od četiri i više točaka (*Roba je pogreškom bila uskladištena u skladištu u*). Osim interpunkcijskih grešaka, QAD također nije pronašao nepostojanje već spomenutih URL adresa iz izvornika u prijevodu, jednu nekonzistentnost izvornika, ponovljene riječi te sve pravopisne greške, zbog nemogućnosti te provjere u alatu. Razlog nepronaska nekonzistentnosti izvornika bilo je korištenje oznake (engl. *tag*) za podebljani tekst u jednoj varijanti u prijevodu (*Ambalaža* i *Ambalaža*), što je alat doživio kao dva različita segmenta te nije javio postojanje dviju verzija u izvorniku (*Packing* i *Packaging material*) za isti termin u prijevodu.

Opciju ignoriranja ovakvih oznaka nude svi alati, unatoč tome što je u spomenutom primjeru vidljivo kako QAD kod ignoriranja oznaka u provjeri konzistentnosti izvornika ne ignorira i oznake u prijevodu. Osim ignoriranja oznaka, alati u nekim provjerama nude i opcije ignoriranja brojeva u riječima (npr. u provjeri pravopisa), riječi pisanih velikim slovima, grešaka koje se javljaju i u izvorniku i tome slično. Sve spomenute opcije mogu biti od koristi u posebnim potrebama provjere u određenom projektu ili zadatku. U ovom istraživanju opcija ignoriranja brojeva unutar riječi u provjeri pravopisa je isključena te su alati Xbench i

Verifika pronašli pravopisnu pogrešku u riječi *prilo2eni* (umjesto *priloženi*), dok memoQ nema dodatne opcije u provjeri pravopisa te nije pronašao spomenutu grešku.

Osim toga, memoQ nema niti opcije provjere URL-a u segmentima i navodnika. Zadnja spomenuta provjera bi se mogla dodati kao posebna provjera za hrvatski jezik, a način na koji je to moguće napraviti je opisan u poglavlju *Dodatne mogućnosti alata*. Provjeru ponovljenih riječi imaju svi alati osim QA Distillera, no jedino je Verifika pronašla grešku kad se radilo o ponovljenoj frazi od dvije riječi (*Ovaj ugovor o zastupanju najprije se najprije se zaključuje na probni rok od šest mjeseci.*), uz onu od ponovljene jedne riječi (*Rezervirali smo vam smještaj koji ste opisal za vrijeme vrijeme koje želite*), koju su pronašli i memoQ i Xbench.

Od svih tipova grešaka uključenih u ovo istraživanje, Xbench nema mogućnosti provjere nedosljednih velikih početnih slova između izvornika i prijevoda, interpunkcijskih znakova i razmaka oko njih ili drugih znakova. U slučaju grešaka dvostrukih razmaka u prijevodu, Xbench nije pronašao one segmente u kojima se dvostruki razmaci pojavljuju i u izvorniku, a u slučaju pogrešnih navodnika našao je samo jednu grešku, u kojoj su dva različita navodnika korištena na početku i na kraju riječi (*Vaša pošiljka ZZZ u skladu je s Vašim željama otpremljena brodom outsiderom "ZZZ".*), no nije pronašao ostale greške pogrešne upotrebe navodnika za hrvatski jezik, jer takve provjere zahtijevaju posebno napisano pravilo. Govoreći o brojevima, svi su se alati pokazali točnima kad se radi o nekonzistentnosti brojeva u izvorniku i prijevodu (npr. *We are looking for an experienced agent for our products as from . . . – Od 2.2. tražimo iskusnog zastupnika za naše proizvode.*), a memoQ, Verifika i QA Distiller imaju i dodatne provjere formatiranja brojeva u skladu s pravilima ciljnog jezika. Za hrvatski jezik vrijedi pravilo pisanja decimalnih brojeva sa zarezom⁴⁸ te odvajanja bojeva većih od 10 000 bjelinom (dopuštena je i točka, no nikako ne i zarez)⁴⁹. Osim u matematici, brojevi ispod 10 000 se ne bi smjeli odvajati bjelinom ili točkom.

Ostale četiri greške, također ubrojene u pogreške brojeva, imale su pogrešno zapisane decimalne brojeve ili tisućice (npr. *Izračun vrijednosti metrike CPU-a: za svaki CPU, prva jezgra procesora množi se s 1, a svaka inkrementalna jezgra procesora množi se s 0.5.*). Iako ovaj tip pogreške naizgled nije bitan toliko koliko je bitna usklađenost među brojevima izvornika i prijevoda, u nekim slučajevima on može činiti značajnu razliku, pogotovo ako se radi o englesko-hrvatskoj jezičnoj kombinaciji. U engleskom jeziku se decimalni brojevi

⁴⁸ Hrvatski pravopis. Mrežno izdanje. Zarez. URL: <http://pravopis.hr/pravilo/zarez/60/> [19.8.2017.]

⁴⁹ Hrvatski pravopis. Mrežno izdanje. Bjelina. URL: <http://pravopis.hr/pravilo/bjelina/54/> [19.8.2017.]

odvajaju točkom, a između stotica i tisućica se piše zarez⁵⁰, suprotno od pravila u hrvatskom jeziku. To može značiti da npr. broj 10,356 napisan u hrvatskom jeziku je zapravo zaokruženo broj 10, a u engleskom bi to bio broj veći od 10 000.

Prilikom pregledavanja pogrešaka određene različitosti među alatima zahtijevale su prilagodbu prebrojavanja grešaka kako bi rezultati bili usklađeni za konačan izračun. Ovo se ponajprije odnosi na različitosti u javljanju jednog tipa pogreške više puta u istom segmentu, što su svi alati osim Verifike brojali kao više grešaka, a ne jednu. To nije vrijedilo za apsolutno svaki tip pogreške u alatima, a memoQ se pokazao alatom najviše osjetljivim na pojedinačno brojanje svake nađene greške u segmentu (npr. u maloprije spomenutom segmentu s brojem, tj. datumom, kojeg nema u izvorniku: *Od 2.2. tražimo iskusnog zastupnika za naše proizvode.*, alat bi javio dvije greške brojeva), što je vidljivo i iz sveukupnog broja nađenih grešaka (1889), koji je daleko veći od onog u svim ostalim alatima.

Za potrebe računanja stupnja pogreške bilo je potrebno na neki način uskladiti broj javljenih stvarnih grešaka među alatima, stoga su se u slučaju stvarnih (t) grešaka gledale greške na razini segmenata unutar svakog tipa. To bi značilo da će se greška pogrešnog početnog i završnog navodnika unutar segmenta brojati kao jedna u tipu greške navodnika (ne kao dvije, što je bio slučaj kod QA-a u QAD-u), no isti će se segment možda pojaviti i unutar nekog drugog tipa greške i tad će se isto brojati. U slučaju pravopisnih grešaka brojale su se sve pojave pogrešno napisane određene riječi, što i ima više smisla od onog kod greške navodnika ili broja/datuma 2.2. Rezultati nađenih grešaka među alatima su se pokušali maksimalno uskladiti, iako je svaka greška u prebrojavanju utjecala i na konačan stupanj pogreške alata, no ponekad je zadatak njihova brojanja bio dug i iscrpan zbog određenih nedostataka u alatu. Već je ranije spomenuto kako Xbench ne omogućuje ignoriranje nekog tipa pogreške u samo jednom segmentu, što je također otežalo konačno brojanje stvarno pronađenih grešaka i njihovo odvajanje od onih koje su krivo nađene. Osim Xbena, niti memoQ nema podatke o detaljnom broju ignoriranih i stvarnih grešaka prema tipovima u svom izvještaju. Verifika i QAD su to riješili na način da u izvještaju postoji podatak o broju ignoriranih i stvarnih grešaka iz grupe grešaka koje svaki alat nudi. Unatoč tome, i te je brojke trebalo uskladiti za potrebe ovog istraživanja, jer svaki alat ima različite grupe provjera koje nudi.

⁵⁰ International Language Environments Guide. Decimal and Thousands Separators. Web page. URL: <https://docs.oracle.com/cd/E19455-01/806-0169/overview-9/index.html> [19.8.2017.]

Najveći problem u prebrojavanju grešaka bile su greške interpunkcijskih znakova, zbog velikog broja pojavljivanja trotočke u prijevodnoj memoriji na kojoj se radilo istraživanje. Pravilo pisanja trotočke u hrvatskom jeziku je uz određenu riječ, s bjelinom nakon trotočke, ili u zagradi (...) ⁵¹, stoga su sve drugačije upotrebe trotočke bile ubrojene u pogrešku interpunkcije (npr. *Oduzeti iznos u visini od ... uplatit ćemo na Vaš račun kod nas.*, a ispravno bi bilo: *Oduzeti iznos u visini od... uplatit ćemo na Vaš račun kod nas.*).

Ako govorimo o važnosti pronalaska određenog tipa pogreške, pogreške bjeline oko interpunkcijskih znakova neće se smatrati toliko bitnima kao npr. pravopisne ili terminološke pogreške, no svejedno bi trebale biti dio osiguranja kvalitete prijevoda, ne samo zbog gramatički ispravnog teksta prijevoda, već i zbog dosljednosti i daljnjeg rada s istim resursom u CAT alatu, kojem će ovakve nesuglasnosti predstavljati problem.

Provjera pravopisa u QA alatima povezana je s provjernikom pravopisa koji je dio nekog drugog alata (MS Word ili Hunspell), o čemu je već bilo riječi u prethodnim poglavljima. Unatoč tome što se alati povezuju na isti provjernik pravopisa, dodatne opcije provjere se razlikuju među alatima. U alatu memoQ je samo moguće odabrati želimo li uključiti pravopisnu provjeru u QA ili ne, što je rezultiralo s najviše krivo nađenih grešaka. Verifika i Xbench nude neke dodatne postavke provjere pravopisa, kao što su ignoriranje ili uključivanje grešaka riječi pisanih velikim slovima, kombinacijom velikih i malih slova, pojave iste riječi u izvorniku, pojave riječi s brojevima i tome slično. Iz tog su razloga oba alata pronašla manji broj krivo nađenih grešaka, u odnosu na memoQ alat.

Još jedna od važnijih opcija koje nude Verifika (ali i memoQ) je mogućnost prijedloga ispravljene pogreške i unošenja automatskog ispravka. Osim toga, Verifika ima i mogućnost dodavanja riječi u rječnik, u slučaju kad se u stvari ne radi o grešci, nego riječ iz nekog razloga nije bila dio rječnika za hrvatski jezik u provjerniku pravopisa za taj ciljni jezik. Dodavanjem riječi u rječnik za taj određeni jezik, ona se više neće pojavljivati u QA izvještaju, stoga je i broj krivo nađenih grešaka Verifike u ovom istraživanju bio najmanji, jer su neke riječi dodane u rječnik i tako isključene iz QA izvještaja (npr. *dogovoreno, generalno, kontrolnog, korisničkog, otkazujem, otvaranjem* itd.).

Provjera terminologije radila se na temelju terminološke baze termina dobivenih ekstrakcijom u alatu memoQ, u zadatku koji je prethodio analizi kvalitete prijevoda (poglavlje 5.3. *Ekstrakcija terminologije*). Provjera ispravne upotrebe termina i njihove dosljednosti od velike je važnosti najviše u prijevodima stručne tematike, jer su tekstovi tog tipa ujedno i

⁵¹ Hrvatski pravopis. Mrežno izdanje. Trotočka. URL: <http://pravopis.hr/pravilo/trotocka/64/> [19.8.2017.]

najbogatiji terminima. U ovome istraživanju dan je primjer važnosti terminološke provjere, unatoč tome što se ne radi o tekstu toliko bogatom terminima, kao što bi to možda bio slučaj s tekstom neke druge tematike. Svi proučavani alati imaju uključenu provjeru terminologije kao dio osiguranja provjere kvalitete prijevoda. Razlike među alatima bile su u formatima i pripremi datoteke s terminima izvornika i prijevoda, kao što je to opisano u prethodnim poglavljima. Broj alatom nađenih grešaka se također razlikovao, ovisno o dodatnim opcijama provjere koje alat nudi (slično kao i kod drugih tipova grešaka, to su uključivanje ili ignoriranje oznaka u tekstu, velikih i malih slova i sl.), no svaki je alat pronašao 42 stvarne terminološke greške u prijevodu. Neke od njih su dane u Tablici 6.

Terminološka baza <i>Poslovno dopisivanje en-hr</i> <i>TB</i>		Prijevodna memorija <i>Poslovno dopisivanje en-hr TM</i>	
eng	hrv	izvornik	prijevod
a buying agent	komisionar za kupnju	Could I also act as a buying agent on your behalf?	Bih li mogao za Vas raditi i kao komisionar za prodaju ?
sales	prodaja	We will be glad that you use our services and favorable sales opportunities .	Bit će nam drago da koristite naše usluge i mogućnosti povoljne kupovine .
agency agreement	ugovor o zastupanju	Termination of agency agreement by the company without notice	Tvrtka trenutavno otkazuje zastupstvo
license agreement	licenčni ugovor	each database product is subject to its respective vendor license agreement ;	svaki proizvod baze podataka podliježe svom odgovarajućem ugovoru o licenci dobavljača;
...

Tablica 6. Primjeri terminoloških grešaka

U prvom i drugom stupcu tablice dani su termini izvornika i prijevoda iz terminološke baze, a u trećem i četvrtom stupcu su segmenti u kojima su pronađene greške za taj par termina. U prvom primjeru je termin *a buying agent* iz izvornika preveden kao *komisionar za prodaju*, umjesto *komisionar za kupnju*. U drugom primjeru nedostaje prijevod termina *sales – prodaja*, a umjesto njega je iskorištena riječ potpuno suprotnog značenja (*mogućnosti povoljne kupovine*, umjesto *mogućnosti povoljne prodaje*). Treći i četvrti primjer pokazuju slučaj kad pogrešna upotreba termina ne predstavlja i značajno značenjsko odstupanje, jer se samo radi o drugoj varijanti prijevoda (*zastupstvo*, umjesto *ugovor o zastupanju* i *ugovor o licenci*, umjesto *licenčni ugovor*), no greške tog tipa važne su za konzistentnost prijevoda. Ako je na više mjesta u dvojezičnoj datoteci *license agreement* preveden s *licenčni ugovor*, a samo na jednom s *ugovor o licenci*, potrebno je uskladiti prijevod tog termina, kako bi se uklonila nedosljednost u terminologiji.

Kako je već ranije spomenuto, sveukupan broj nađenih terminoloških grešaka se razlikovao u alatima, a u alatu Verifika je on bio najmanji (254), jer on jedini ima opciju dodavanja oblika za termine (slično kao i dodavanje riječi u rječnik prilikom pregledavanja pravopisnih grešaka). Unatoč tome što i Verifika pronalazi grešku npr. *inland waterway – unutarnja plovidba* u segmentu: *Paleta usluga koju nudite toliko je uvjerljiva da Vas ovime želimo angažirati za sve naše prijevoze **unutarnjom plovidbom.***, dodavanjem tog oblika za termin unutarnja plovidba, ta se greška uklanja s popisa i iz QA izvještaja. Osim toga, svaka sljedeća pojava termina u tom padežu (također i u novim projektima), neće biti dio nađenih grešaka prilikom provjere kvalitete prijevoda. Iz tog se razloga ukupan broj nađenih grešaka u ovom alatu smanjio i odstupa od ukupnog broja terminoloških grešaka koje su pronašli ostali alati.

5.4.5. Dodatne mogućnosti alata

Cilj usporedbe alata bio je ukazati na njihove sličnosti i razlike u procesu provjere kvalitete prijevoda s unaprijed zadanim postavkama za ciljni jezik ili malim intervencijama u zadanim postavkama za što precizniju usporedbu. To je ujedno isključilo neke dodatne mogućnosti alata koje bi mogle utjecati na njihov stupanj pogreške, stoga će u ovom dijelu rada biti detaljnije opisane neke od značajki alata koje nisu dio zadanih postavki te zahtijevaju manje ili veće znanje rada s proučavanjem alatima te regularnim izrazima.

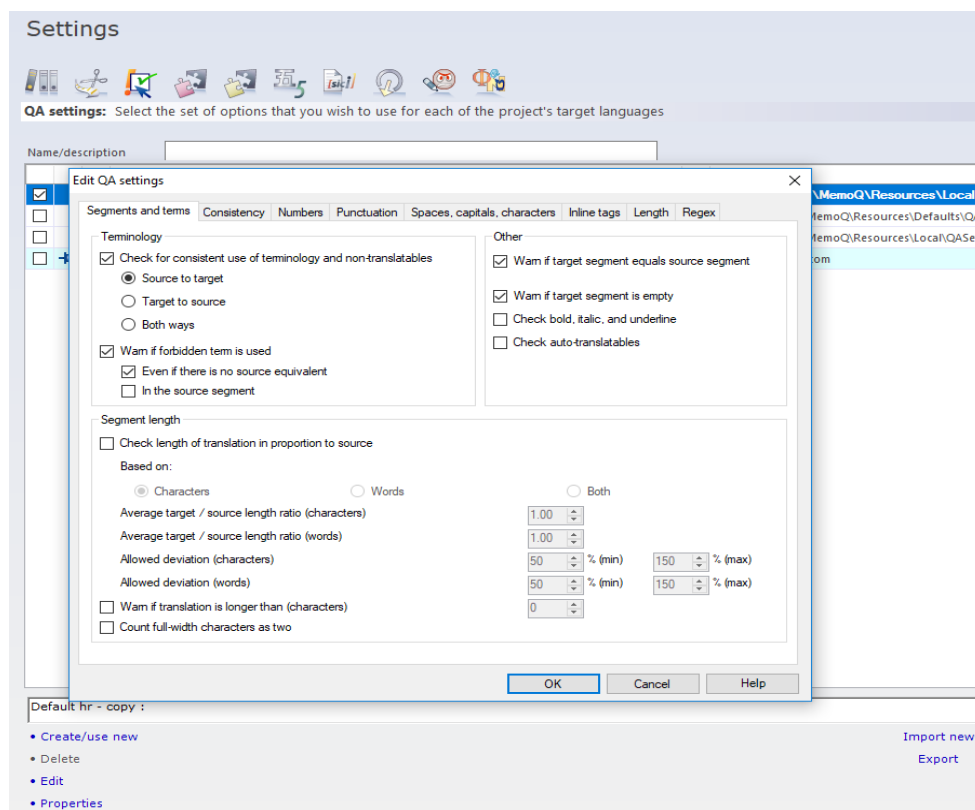
5.4.5.1 Profili

Broj jezika koje alati podržavaju varira od jednog do drugog alata, no većina alata ima već unaprijed zadana pravila za jezike koje podržava. Ta se pravila tiču npr. interpunkcije, pisanja decimalnih brojeva u jeziku, navodnika i tome slično. Iz razloga što često nije poznat izvor unaprijed definiranih pravila za jezike, njihova je točnost također upitna. Uređivanje postavki za jezike zahtijeva veće znanje o alatima te višu razinu tehničke pismenosti (ukoliko se, primjerice, odlučimo za dodavanje nekog pravila u obliku regularnog izraza), stoga je od velike važnosti mogućnost spremanja profila u alatima. Svrha stvaranja profila za pojedine jezike je njihovo ponovno pokretanje kod novog prijevoda na isti ciljni jezik. To će ujedno omogućiti i konzistentno ispravljanje grešaka koje će se i u budućnosti javljati u prijevodima na isti jezik.

Stvaranje profila razlikuje se od alata do alata pa tako unutar memoQ alata to znači uređivanje QA postavki te njihovo spremanje i ponovno pokretanje u određenom projektu.

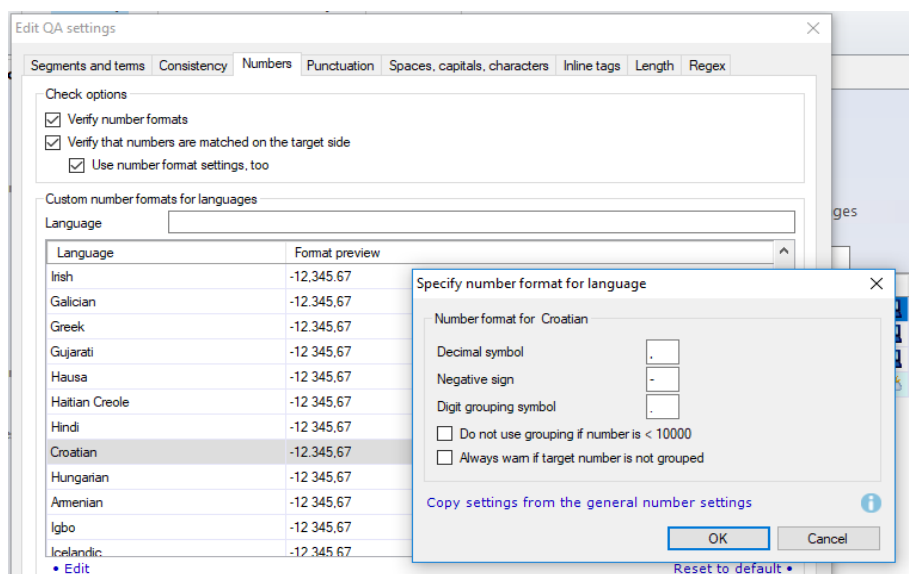
Mogućnosti uređivanja postavki memoQ-a su općenite provjere segmenata i termina, dosljednosti, brojeva, interpunkcije, razmaka i slova, tagova, duljine te posebne provjere koje

je moguće dodati u obliku regularnih izraza. Nakon namještanja svih postavki za, primjerice, hrvatski jezik, one mogu biti spremljene pod odabranim nazivom te željenom oznakom za hrvatski jezik i eksportane u *.mqres formatu. Osim eksporta, moguće je i napraviti import datoteke istog formata sa spremljenim postavkama za neki jezik ili za potrebe određenog projekta/klijenta.



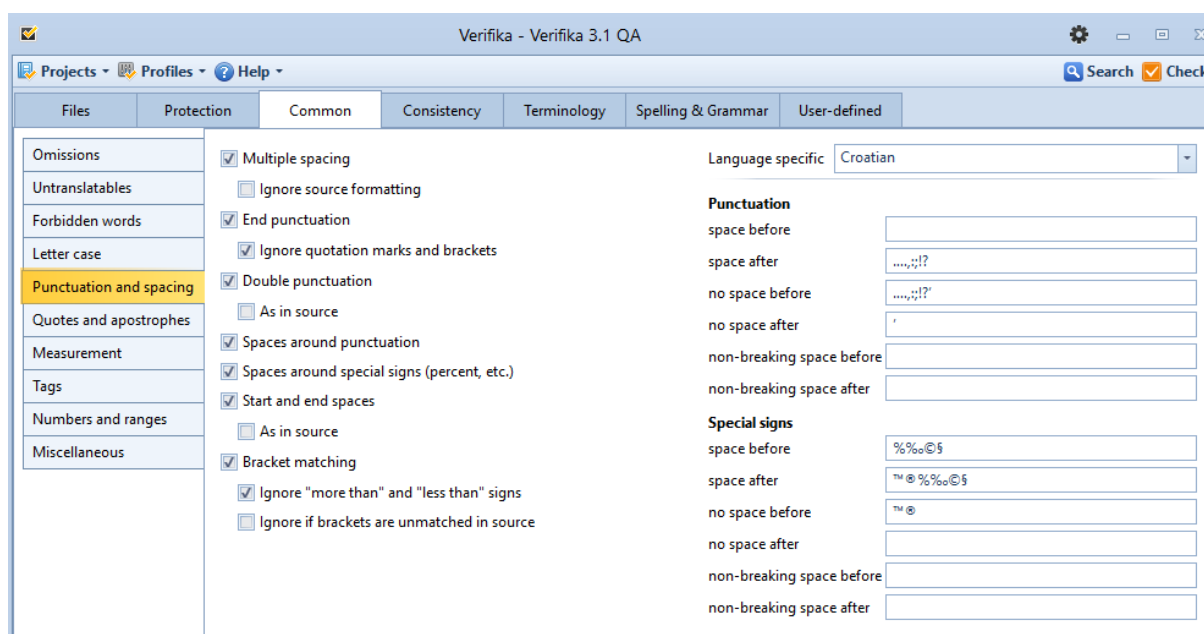
Slika 23. Postavke provjere kvalitete – memoQ

Sve provjere osim brojeva i interpunkcije dio su nespecifičnih postavki i iste su za sve jezike ako ih posebno ne prilagodimo željenim potrebama. Za brojeve i interpunkciju moguće je odabrati već postojeća pravila za traženi jezik, a i dodatno ih urediti te kao takve spremiti. Slika 24. prikazuje uređivanje postavki za pravila pisanja decimalnih brojeva i tisućica u hrvatskom jeziku.



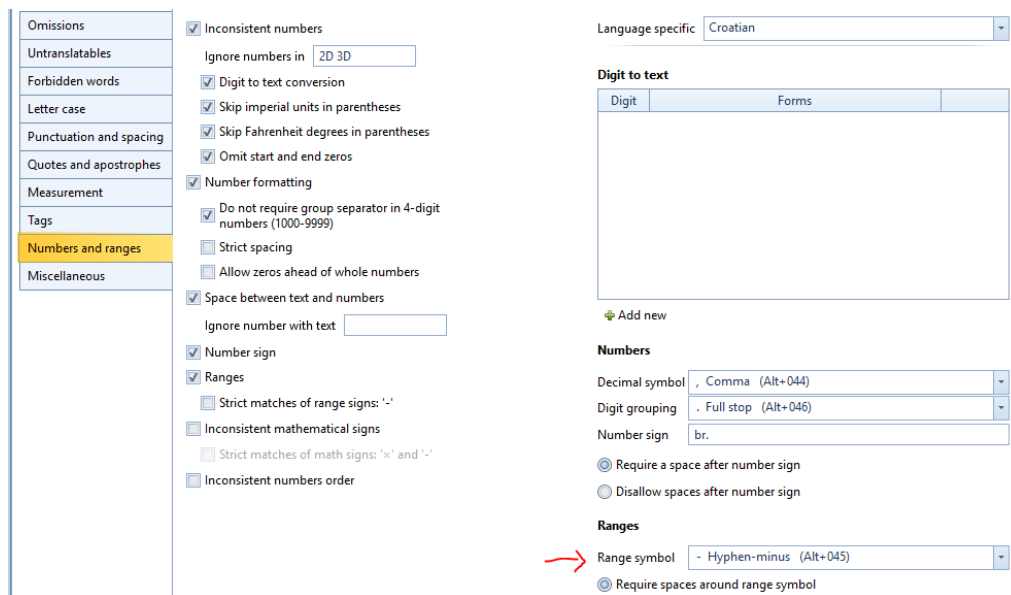
Slika 24. Uređivanje postavki za pisanje brojeva u hrvatskom jeziku – memoQ

Iako je stvaranje profila u alatu Verifika s jedne strane vrlo slično onom opisanom u memoQ-u, ono je ipak više prilagođeno jezicima prijevoda. Prilikom ubacivanja dvojezične datoteke u sam alat, automatski se učitavaju programom zadana pravila za ciljni jezik. Jezik koji Verifika automatski prepoznaje je jezik određen CAT alatom prilikom kreiranja projekta. Na Slici 25. su vidljiva unaprijed određena pravila i namješten profil za hrvatski koji se učitao prilikom ubacivanja dvojezične en-hr datoteke na kojoj se radio praktični dio ovog rada.



Slika 25. Uređivanje postavki profila za hrvatski jezik – Verifika

Unatoč tome što unaprijed određena pravila za jezike u većini odgovaraju stvarnom pravopisu tog jezika, to nije tako u potpuno svakom slučaju. Slika 26. prikazuje jedan primjer pogrešnog pravila za hrvatski, a tiče se pisanja raspona između brojeva. Prema pravopisu hrvatskog jezika ispravno pisanje raspona je korištenjem crtice – (Alt + 0150), kao u primjerima: *2010. – 2012.*, *autocesta Split – Zagreb*, *let New York – London*, *ponedjeljak – petak*, *30 – 45 minuta*, *A – Ž*.⁵²



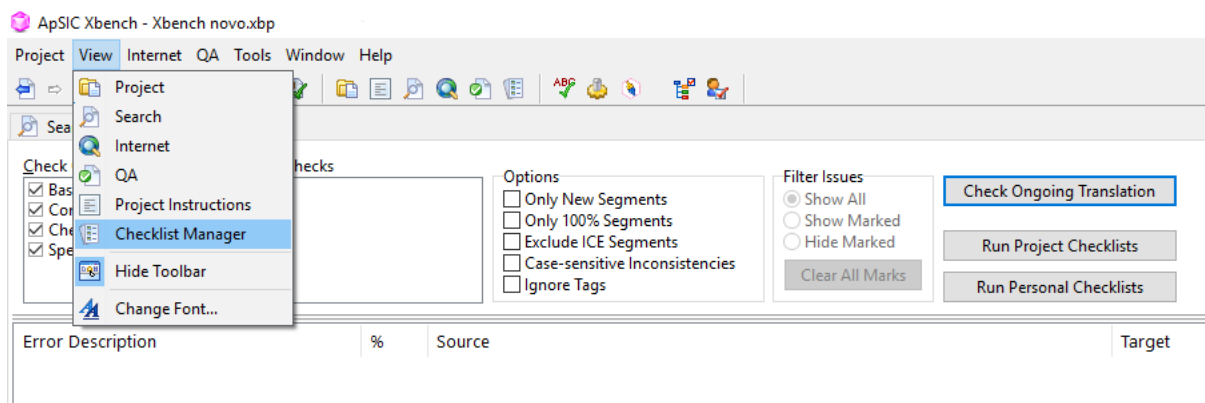
Slika 26. Pogrešno pravilo pisanja raspona brojeva (Verifika)

Iz tog razloga je potrebno promijeniti zadano pravilo te spremi novi profil za hrvatski jezik i učitati ga prilikom provjere kvalitete novog prijevoda na hrvatski.⁵³

U alatu Xbench profili se nazivaju listama provjere (engl. *Checklists*), a pravila se u njih dodaju ručno upisivanjem naziva, opisa i vrste provjere te traženog simbola, riječi ili niza riječi u izvorniku i/ili ciljnom jeziku. Pretraživati je moguće i koristeći obične znakove, no svako složenije pravilo zahtijeva poznavanje naprednijeg pretraživanja pomoću MS Word Wildcarda ili regularnih izraza.

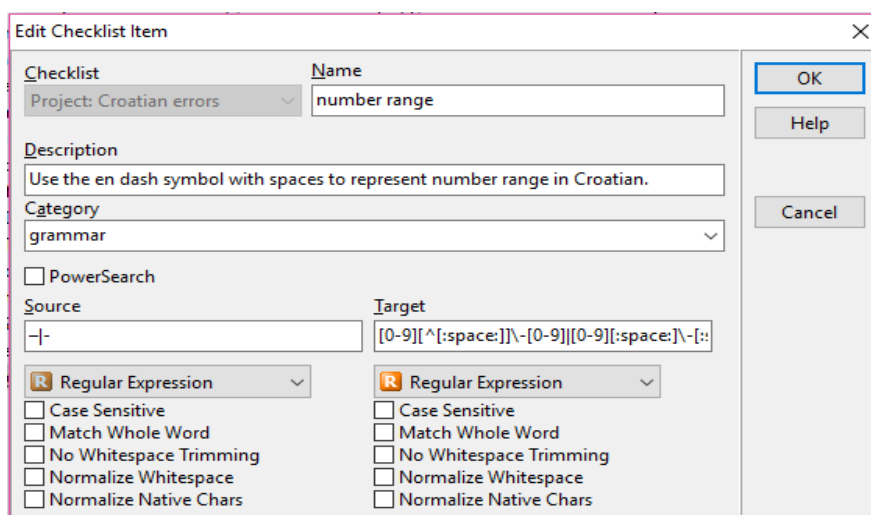
⁵² Hrvatski pravopis. Mrežno izdanje. Crtica. URL: <http://pravopis.hr/pravilo/crtica/69/> [10.9.2017.]

⁵³ Spremanje profila (Verifika): Profiles – Save – upisivanje željenog naziva
Učitavanje postojećeg profila (Verifika): Profiles – Load – odabir datoteke



Slika 27. Upravljanje listama provjere (Xbench)

Na Slici 28. prikazan je primjer dodavanja provjere pisanja raspona brojeva u hrvatskom jeziku.

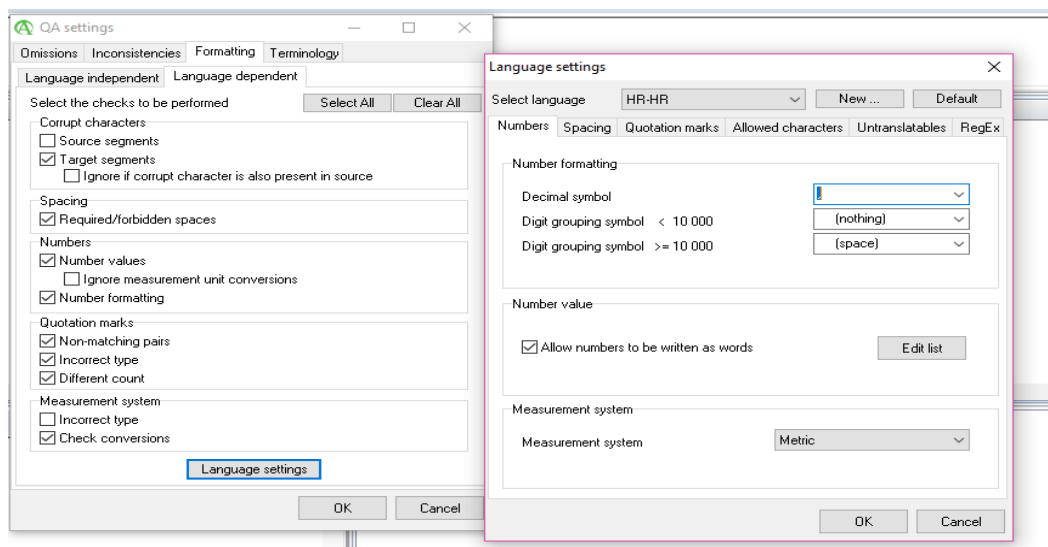


Slika 28. Dodavanje novog pravila u listu provjere (Xbench)

U polju za unos simbola izvornika upisani su - ili –, a u polju za unos simbola ciljnog jezika upisan je duži regularni izraz koji pronalazi i upozorava na sve pogrešne upotrebe pisanja raspona brojeva u hrvatskom jeziku (npr. 5-9 ili 5 - 9 ili 5–9).

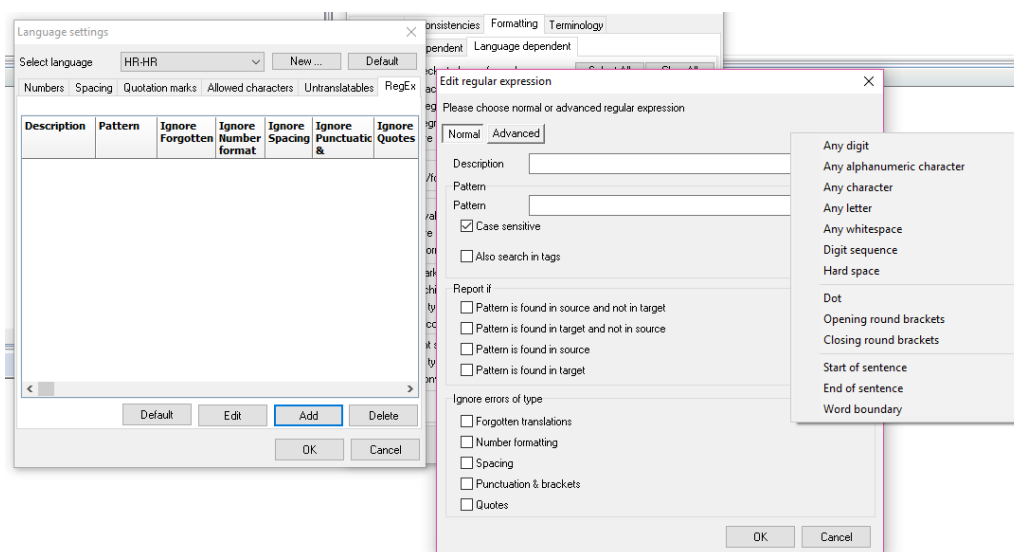
Iako pisanje pravila u Xbench-u zahtijeva višu razinu tehničke pismenosti, njima je moguće dodati vrlo detaljne provjere kvalitete prijevoda ili posebnih zahtjeva nekog projekta, odnosno klijenta.

Kreiranje profila jedna je od značajki i alata QA Distiller. U ovom je alatu kreiranje profila osmišljeno kao uređivanje QA postavki (Slika 29.), onih koje nisu jezično ovisne i onih koje jesu jezično ovisne.



Slika 29. Uređivanje profila (QA Distiller)

U dodatnim postavkama za jezike moguće je upisati pravila u obliku regularnih izraza, kao što je prikazano na sljedećoj slici.

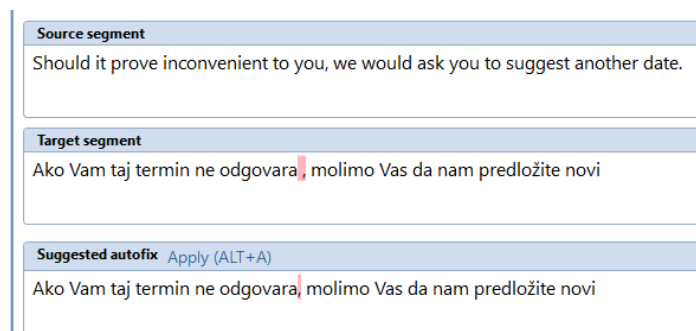


Slika 30. Korištenje regularnih izraza u profilu (QA Distiller)

Sve napravljene izmjene u postavkama spremaju se u mapu *userlists*, koja je dio instalacijskog paketa programa. Njezinim kopiranjem na novu lokaciju i preimenovanjem mogu se dobiti gotove postavke za određeni jezik ili projekt. Datoteke s postavkama u mapi *userlists* se spremaju u *.ini ili *.txt formatu datoteke, prema različitim vrstama postavki i pretraživanja, a mogu se uređivati, osim u samom alatu, i u bloku za pisanje.

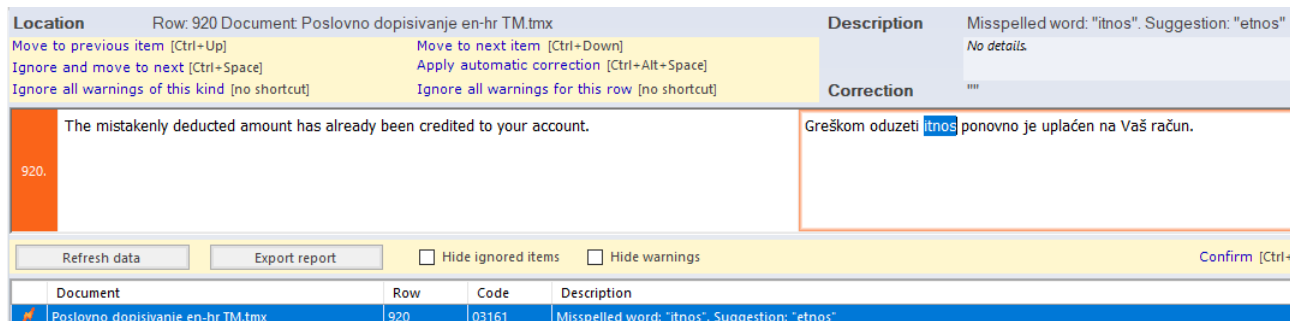
5.4.5.2 Samoispravak (engl. Autocorrection)

Dodatna mogućnost koju nude neki alati je prijedlog za ispravak nađene greške (engl. *suggested autofix*). Ona može biti od velike koristi pogotovo kad se radi o greškama formatiranja, bjelinama i sl. (Slika 31.), jer omogućuje vrlo jednostavnu ispravku greške u ciljnom segmentu (korištenje prečaca Alt+A na tipkovnici u alatu Verifika).



Slika 31. Prijedlog ispravka nađene greške (Verifika)

U greškama koje su jezično specifičnije prijedlozi ispravka nisu uvijek točni, a to ovisi o tome koliko su u alatu uređene postavke za određeni jezik i koliko je njegovo poznavanje tog jezika. Primjerice, u pravopisnim greškama, često će se raditi o krivom prijedlogu riječi (Slika 32.) pa ova mogućnost neće biti od velike koristi.



Slika 32. Pogrešan prijedlog ispravka (itnos u etnos, umjesto iznos) (memoQ)

Od alata proučavanih u ovom radu jedino Verifika i memoQ imaju opciju prijedloga ispravka pogreške i automatskog unošenja izmjene.

5.4.5.3 Nprevedive riječi (engl. Untranslatables)

Nprevedive (engl. *untranslatables*) riječi česta su pojava u stručnim tekstovima, a najčešće se odnose na nazive brandova (npr. **Siemens**, **Bosch**, **Samsung**), uređaja (npr. ... uporabu sa sljedećim popularnim merijima digitalne kamere: **Memory Stick™**, **Memory Stick PRO™**,

*Memory Stick Duo™, Memory Stick PRO Duo™, SD i SDHC.*⁵⁴), riječi stranog porijekla (npr. *schadenfreude* u engleskom jeziku) ili riječi iz uputa u korisničkim priručnicima koje ostaju na engleskom jeziku (npr. *zatim kliknite Copy to local disk*⁵⁵). U radu na prevoditeljskom projektu, popis neprevodivih riječi dio je posebnih uputa klijenta za taj projekt, a u procesu provjere kvalitete, kao završnom stupnju provjere u projektu, provjerava se jesu li se prevoditelji/lektori držali danih uputa. Mogućnost dodavanja liste neprevodivih riječi u provjeru kvalitete imaju svi proučavani QA alati, iako se u Xbench-u ta provjera može napraviti jedino na način da se u projekt doda lista termina s jednakim stupcem izvornika i prijevoda. Ostali alati imaju više funkcija dodavanja i uređivanja postavki za neprevodive riječi.

5.4.5.4 Zabranjene riječi (engl. *Forbidden words*)

Slično poput liste neprevodivih riječi, QA alati nude i mogućnost uvrštavanja liste zabranjenih riječi u sklopu provjere kvalitete. Zabranjene riječi su riječi koje se ne smiju ili ih nije preporučljivo koristiti u prijevodu, a najčešće ih određuje sam klijent. QA alati služe za provjeru korištenja zabranjenih riječi iz popisa te njihova uklanjanja prema potrebi. Zadatak kontrolora kvalitete ili lektora za ciljni jezik je preoblikovati željenu misao korištenjem riječi ili fraze sličnog značenja, kako bi se izbjegla ona riječ za koju u uputi stoji da se ne koristi.

⁵⁴ Osnovni korisnički priručnik. 2010 Bother Industries, Ltd. URL: http://download.brother.com/welcome/doc002692/cv_dcp315w_cro_busr.pdf 31. str. [10.9.2017.]

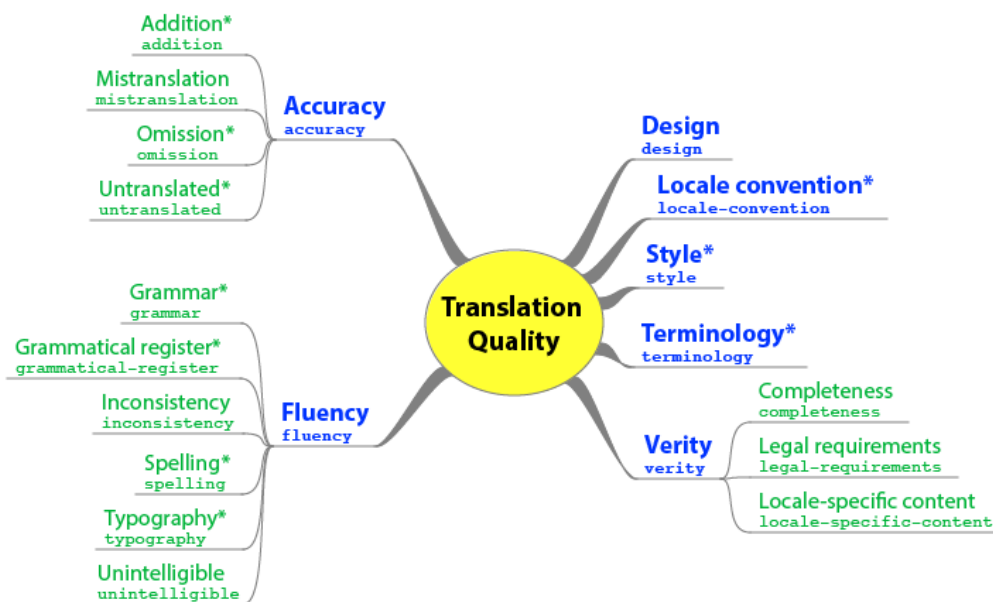
⁵⁵ Osnovni korisnički priručnik. 2010 Bother Industries, Ltd. URL: http://download.brother.com/welcome/doc002692/cv_dcp315w_cro_busr.pdf 10. str. [10.9.2017.]

5.5. Evaluacija kvalitete prijevoda

Zadnji zadatak u ovom istraživanju je bio računanje kvalitete prijevoda na temelju grešaka pronađenih u prethodnom zadatku. U nastavku će biti opisana metoda koja je korištena za evaluaciju kvalitete, a nakon toga i sama evaluacija.

5.5.1. Višedimenzionalna metrika za evaluaciju kvalitete

Višedimenzionalna metrika za evaluaciju kvalitete (engl. *Multidimensional Quality Metrics* (MQM)) je model koji služi za opis kvalitete prijevoda prema definiranim kategorijama problema.⁵⁶ MQM definira više od 100 tipova problema koji su opisani hijerarhijski. Takav opis problema koje sustav proučava omogućava korištenje metrike na bilo kojem stupnju hijerarhije. Primjerice, moguće je promatrati samo probleme najvišeg hijerarhijskog stupnja: točnost (engl. *Accuracy*) i fluentnost (engl. *Fluency*), ali i proširiti popis problema na njihove podskupine ili druge skupine problema.⁵⁷ Ovisno o potrebama određenog projekta, koriste se različiti stupnjevi problema u evaluaciji kvalitete. Jezgru MQM sustava čini 20 tipova problema koji predstavljaju najčešće vrste problema u procjeni kvalitete prijevoda. Na Slici 33. se nalazi grafički prikaz 20 osnovnih tipova problema u MQM sustavu.



Slika 33. Jezgra MQM-a

⁵⁶ Multidimensional Quality Metrics. 2014. URL: <http://www.qt21.eu/launchpad/content/multidimensional-quality-metrics> [20.8.2017.]

⁵⁷ Multidimensional Quality Metrics (MQM) Issue Types. German Research Center for Artificial Intelligence (DFKI) and QTLaunchPad. 2015. URL: <http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html> [20.8.2017.]

Problemi najvišeg stupnja u hijerarhiji su već spomenute točnost i fluentnost, a popis se nastavlja na dizajn (engl. *Design*), probleme lokalne konvencije (engl. *Locale convention*), stil (engl. *Style*), terminologiju (engl. *Terminology*), istinitost (engl. *Verity*). Podskupine problema točnosti su dodaci prijevodu, pogrešan prijevod, izostavljen ili nepreveden tekst, a podskupine problema fluentnosti su gramatika, konzistentnost, pravopis, tipografija i nečitljivost.

5.5.1.1. Vrednovanje

Neobavezni dio MQM metrike je vrednovanje problema prema težini. Zadana težina svakog problema je 1. Prema posebnim zahtjevima nekog zadatka ili projekta, problemi mogu dobiti veću mjeru težine⁵⁸, i to na nekoliko razina:

- problemi bez težine (engl. *none*, 0) – oni problemi koji nisu važni za prijevod i koje nije potrebno riješiti, a služe samo kao upozorenje ili napomena u vezi projekta ili rada na budućim projektima;
- manji problemi (engl. *minor*, 1) – oni problemi koji ne utječu na sam sadržaj prijevoda, a tiču se grešaka poput pogrešnih i duplih bjelina u tekstu i tome slično;
- veći problemi (engl. *major*, 5) – oni problemi koji utječu na sadržaj prijevoda, ali ga i dalje ne čine nerazumljivim, npr. pravopisna pogreška će zahtijevati ispravljanje, ali je značenje prijevoda moguće shvatiti i bez njenog ispravljanja;
- kritični problemi (engl. *critical*, 10) – oni problemi koji utječu na značenje prijevoda (npr. ozbiljnije gramatičke pogreške, izostavljeni prijevod itd.).⁵⁹

5.5.2. Evaluacija kvalitete

Prethodno opisana metoda evaluacije kvalitete prijevoda je odabrana kao metrika za ručno vrednovanje kvalitete prijevoda u prijevodnoj memoriji *Poslovno dopisivanje en-hr TM*. Stvarne greške koje su pronašli QA alati bit će podijeljene u skupine problema prema onima koje su opisane u višedimenzionalnoj metrici za evaluaciju kvalitete prijevoda.

⁵⁸ Mjere za težinu problema 0, 1, 5 i 10 preuzete su iz:

Yang, J.; Ciobanu, D.; Reiss, C; Secară, A. Using Computer Assisted Translation Tools' Translation Quality Assesment functionalities to assess students' translations. *The Language Scholar Journal*. 2017. URL: <http://languagescholar.leeds.ac.uk/wp-content/uploads/sites/3/2017/05/Issue-1-YangReissSecaraCiobanu.pdf> [19.9.2017.] Str. 13

⁵⁹ Multidimensional Quality Metrics (MQM) Issue Types. Scoring. German Research Center for Artificial Intelligence (DFKI) and QTLanchPad. 2015. <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html> [20.8.2017.]

Osim podjele grešaka prema skupinama iz MQM-a, one će biti vrednovane i prema razinama težine problema (mjera 1 za manje probleme, 5 za veće probleme i 10 za kritične probleme). U Tablici 7. je 12 prethodno korištenih tipova grešaka svrstano u skupine problema iz MQM-a i svakom je tipu greške dodijeljena mjera težine.

	TIP POGREŠKE	Broj grešaka (t _{max})	Težina
Točnost	Terminologija	42	5
	Brojevi	9	10
	Ponovljene riječi	2	10
	Pogreške URL-a	2	10
Fluentnost	Pravopis	53	5
	Konzistentnost prijevoda	24	5
	Konzistentnost izvornika	10	5
	Veliko slovo	1	1
	Dvostruki razmaci	29	1
	Interpunkcijske pogreške	109	5
	Navodnici	47	1
	Razmaci oko znakova	204	1

Tablica 7. Rezultati QA alata za evaluaciju kvalitete prijevoda

Terminološke pogreške, pogreške brojeva, ponovljenih riječi i URL-a su svrstane u skupinu točnosti, a ostale greške pripadaju skupini fluentnosti. Težine problema su dodijeljene prema prethodno definiranoj podjeli važnosti problema za razumijevanje značenja teksta. Najveću težinu su dobile pogreške brojeva, ponovljenih riječi i URL-a, a najmanju pogreške velikog slova, dvostrukih razmaka, navodnika i razmaka oko znakova. Za vrijednosti broja grešaka uzete su vrijednosti t_{max} iz prethodnog zadatka (Tablica 5.).

Ocjena kvalitete

Ocjena kvalitete (engl. *Translation Quality*) bit će izračunata prema formuli:⁶⁰

$$TQ = 100 - AP - (FP_T - FP_S);$$

gdje su:

TQ – ocjena kvalitete,

⁶⁰ Lommel, Arle Richard, Burchardt, Aljoscha, Uszkoreit, Hans. Multidimensional Quality Metrics; A flexible System for Assessing Translation Quality, Berlin. 2013. URL: <http://www.mt-archive.info/10/Aslib-2013-Lommel.pdf> [10.9.2017.] Str. 6

AP – kaznena vrijednost pogrešaka iz skupine točnosti,

FP_T – kaznena vrijednost pogrešaka iz skupine fluentnosti za ciljni tekst,

FP_S – kaznena vrijednost pogrešaka iz skupine fluentnosti za izvorni tekst (izvorni tekst se ne ocjenjuje, dakle FP_S = 0).⁶¹

Kaznena vrijednost (engl. *Penalty*) za sve tipove problema računa se prema formuli:

$$P = (I_{\text{minor}} + I_{\text{major}} \times 5 + I_{\text{critical}} \times 10) / W_c;$$

gdje su:

P – kaznena vrijednost problema,

I_{minor} – problemi manje važnosti,

I_{major} – problemi veće važnosti,

I_{critical} – kritični problemi,

W_c, – broj riječi.⁶²

Uobičajen je zapis kaznenih vrijednosti P u postocima.⁶³ Nakon izračuna kaznenih vrijednosti problema P dobivene su vrijednosti AP = 1,24 % za pogreške točnosti i FP = 4,61 % za pogreške fluentnosti.

Konačna ocjena kvalitete (TQ) iznosi:

$$TQ = 100 - AP - FP;$$

$$TQ = 100 \% - 1,24 \% - 4,61 \%;$$

$$TQ = 94,15 \%$$

⁶¹ Lommel, Arle Richard, Burchardt, Aljoscha, Uszkoreit, Hans. Multidimensional Quality Metrics; A flexible System for Assessing Translation Quality, Berlin. 2013. URL: <http://www.mt-archive.info/10/Aslib-2013-Lommel.pdf> [10.9.2017.] (Iz izvorne formule izbačeni su problemi istinitosti (engl. *verity*))

⁶² Broj riječi ovdje se odnosi na broj riječi ciljnog teksta, jer su u njemu proučavane pogreške. (W_c = 27 354)

⁶³ Generating a translation quality score with MQM. Web page. 2014. URL: <http://www.qt21.eu/launchpad/node/1332> [10.9.2017.]

5.5.3. Rasprava

Konačna je ocjena kvalitete prijevodne memorije korištene u ovom radu, dobivena na temelju grešaka koje su pronašli analizirani QA alati, 94,15 %. Na vrijednost ocjene kvalitete značajno utječu dva važna podatka, a to su ukupan broj riječi i količina stvarnih grešaka. U ovom je istraživanju ukupan broj riječi (W_c) 27 354, a ukupan broj stvarnih grešaka (t_{max}) iznosi 532. Ispravljanjem svih pronađenih grešaka će se kvaliteta prijevoda u prijevodnoj memoriji poboljšati za 5,85 %. Unatoč tome što neke od grešaka nisu bile od kritične važnosti za razumijevanje značenja prijevoda, nužno ih je ispraviti, jer se njihovom uporabom narušava kvaliteta svakog novog prijevoda na koji će biti priključena ova prijevodna memorija. Rezultat ispravljanja svih pronađenih grešaka je točniji resurs više kvalitete od onog prije obavljanja QA-a. Postojeći alati za provjeru kvalitete olakšavaju sam pronalazak grešaka, ali i njihovo ispravljanje. Ljudskom provjerom kvalitete bi pronalaženje svih 532 grešaka zahtijevalo puno više vremena i teže uočavanje tehničkih grešaka i grešaka formatiranja, koje su također dio teksta prijevoda. Alati nude automatski pronalazak grešaka prema namještenim postavkama provjere kojima je moguće na vrlo jednostavan način upravljati, ovisno o tome koje nas provjere zanimaju. Ispravljanje grešaka je više ili manje jednostavno, ovisno o načinu ispravljanja grešaka kakav alat podržava. S obzirom na to da se QA alati još razvijaju i da važnost osiguranja kvalitete prijevoda raste s potrebama i zahtjevima jezičnih projekata, zasigurno će se razvijati i njihove značajke u vidu poboljšanja samog korisničkog sučelja, kao i uključivanja nekih dodatnih opcija, poput mogućnosti samoispravka za veći broj provjera i u više alata. Osim toga, napredak bi mogao ići i u smjeru razvijanja boljih terminoloških i pravopisnih provjera za flektivne jezike, zbog izrazito velikog broja zanemarivih grešaka u QA izvještaju, što ujedno povećava mogućnost preskakanja onih grešaka koje stvarno treba ispraviti. Ujedno bi se trebalo raditi i na općenitom smanjenju broja nađenih zanemarivih grešaka, što zahtijeva detaljnije postavke provjera. Unatoč tome što će gotovo svako pravilo imati neku iznimku i što će neke greške moći biti ignorirane, pronalazak velikog broja stvarnih pogrešaka ukazuje na isplativost korištenja QA alata.

6. Zaključak

Cilj je ovog diplomskog rada bio prikaz korištenja alata i resursa u računalno potpomognutom prevođenju, s posebnim naglaskom na proces osiguranja kvalitete prijevoda (QA). Razvoj računalne tehnologije je rezultirao novim zadacima i poslovima u jezičnoj industriji, kao što su osiguranje kvalitete, upravljanje projektima, analiza podataka te tehnički i programerski zadaci. Osim navedenoga, korištenje tehnologije postavlja nove zahtjeve za obrazovanjem prevoditelja na razini korištenja pojedinog alata, dok obrazovanje informacijskih i komunikacijskih stručnjaka zahtijeva sveobuhvatnije i dublje poznavanje tehnologije, alata i procesa.

Porast količine posla doveo je do narušavanja njegove kvalitete i potrebe za razvijanjem alata koji će omogućiti provjeru i upravljanje projektom. Pogreške koje nastaju prilikom prevođenja rezultat su ljudskog faktora. U slučaju većih projekata javljaju se različite vrste pogrešaka (npr. nedosljednost u korištenju terminologije, korištenje zabranjenih termina, neprevedivih riječi, korištenje dodatnih resursa i uputa, formatiranja, pravopisa, itd.). Alatima za provjeru kvalitete prijevoda je moguće otkriti različite tipove pogrešaka te na razini projekta provesti korekcije i analizu kvalitete prijevoda.

Praktični je dio ovog rada bio usmjeren prema prikazu korištenja alata u računalno potpomognutom prevođenju, s posebnim naglaskom na kvalitetu prijevoda. Prvi je primjer upotrebe alata bio za dobivanje resursa koji su se koristili u zadatku provjere kvalitete. Drugi je primjer upotrebe alata dao uvid u njihove mogućnosti u svrhu poboljšanja kvalitete, dok je analiza alata i pogrešaka pokazala koliko se kvaliteta prijevoda poboljšava njihovom upotrebom i ispravljanjem. Rezultati analize grešaka i računanje ocjene kvalitete prijevoda su potvrdili važnost osiguranja kvalitete kao sastavnog dijela projekta i obrade dokumenta. Osim važnosti osiguranja kvalitete prijevoda u prevedenim dokumentima, u ovom je radu dan primjer provjere kvalitete na prijevodnoj memoriji u kojoj je, zbog utjecaja na kvalitetu u daljnjim projektima, potrebno osigurati visoku kvalitetu njezinog sadržaja, dosljednost i gramatičku ispravnost segmenata.

7. Literatura

1. Alcina, A., Translation Technologies: Scope, tools and Resources, International Journal on Translation Studies. Španjolska. 2008. URL: <http://repositori.uji.es/xmlui/bitstream/handle/10234/18016/28696.pdf> [5.9.2017.]
2. Bowker, Lynne, Fisher, Des. Computer-aided translation. 2010. University of Ottawa.
3. Czopik, Jerzy. Quality Assurance process in translation. 2014. <http://www.mt-archive.info/10/Asling-2014-Czopik.pdf> [10.9.2017.]
4. Dovedan, Zdravko; Seljan, Sanja; Vučković, Kristina. Strojno prevođenje kao pomoć u procesu komunikacije. Str. 283-291. Informatologia 35 (4), 2002. URL: <http://dzs.ffzg.unizg.hr/text/spuk.pdf> [10.9.2017.]
5. Esselink, Bert. A Practical Guide to Localization. John Benjamins Publishing Company. Amsterdam/Philadelphia. 2000.
6. Esselink, Bert. The Evolution of Localization. Solution Architect, Lionbridge. 2000. URL: http://www.intercultural.urv.cat/media/upload/domain_317/arxiu/Technology/Esselink_Evolution.pdf [10.9.2017.]
7. Gow, Francie. Metrics for Evaluating Translation Memory Software. 2003. University of Ottawa, Ottawa, Canada. URL: http://www.chandos.ca/Metrics_for_Evaluating_Translation_Memory_Software.pdf [10.9.2017.]
8. Gural, Svetlana K.; Chemezov, Yan R. Analysis of Efficiency of Translation Quality Assurance Tools / Procedia - Social and Behavioral Sciences 154 (2014) 360 – 363 URL: <http://www.sciencedirect.com/science/article/pii/S1877042814056134> [10.9.2017.]
9. Hartley, Tony. Technology and Translation. // The Routledge Companion to Translation Studies. London; New York : Routledge, 2009. Str. 106-128. URL: http://www.data.ulis.vnu.edu.vn/jspui/bitstream/123456789/2910/1/UTF-8_0415396417Translation.pdf [10.9.2017.]
10. House, Juliane. Translation quality assessment: past and present. New York, 2015. <http://translationindustry.ir/Uploads/Pdf/Translation%20Quality%20Assessment.pdf> [10.9.2017.]

11. Hrvatska enciklopedija. Mrežno izdanje. Leksikografski zavod Miroslav Krleža. Zagreb. 1999–2009. URL: <http://www.enciklopedija.hr/Default.aspx> [10.9.2017.]
12. Hutchins, John. Machine translation: past, present, future. New York: Halsted Press, 1986. URL: <http://www.hutchinsweb.me.uk/PPF-TOC.htm> [10.9.2017.]
13. Hutchins, John. Machine translation: A brief history. 1995. URL: <http://www.hutchinsweb.me.uk/ConcHistoryLangSci-1995.pdf> [10.9.2017.]
14. Hutchins, John. The origins of the translator's workstation. 1998. University of East Anglia, Norwich, UK. URL: <https://pdfs.semanticscholar.org/0bb2/fe15d30b1243f2fc90bed2ecc5a23d57ccfc.pdf> [10.9.2017.]
15. Hutchins, John. The development and Use of Machine Translation Systems and Computer-based Translation Tools. 2003. University of East Anglia, Norwich, UK. URL: <http://mt-archive.info/IJT-2003-Hutchins.pdf> [10.9.2017.]
16. Hutchins, John. Current commercial machine translation systems and computer-based translation tools: system types and their uses. 2005. URL: <http://www.hutchinsweb.me.uk/IJT-2005.pdf> [10.9.2017.]
17. Kockaert, J. Hendrick, Steurs, Frieda, Makoushina, Julia. Translation quality assurance, what is missing? And what can be done? 2008. URL: http://www.ling.arts.kuleuven.be/qlvl/prints/kockaert_steurs_makoushina_2008draft_Translation_quality_assurance.pdf [10.9.2017.]
18. Kostiainen, Tuomas. Tools quality assurance and translation memory maintenance. 25.10.2012. URL: http://www.finntranslations.com/wordpress/wp-content/uploads/2010/08/LT-2_handout_TuomasKostiainen.pdf [10.9.2017.]
19. Kučiš, Vlasta. Prevodilački alati u funkciji kvalitete prijevoda. 2009. Filozofski fakultet, Sveučilište u Mariboru, Maribor, Slovenija. Str. 19 – 33. URL: <http://hrcak.srce.hr/53039> [10.9.2017.]
20. Le Ny, Benoit. Quality Control in Human Translations: Use Cases and Specifications. 2014. URL: https://transread.limsi.fr/Deliverables/Deliverable_3.1.pdf [10.9.2017.]

21. Lommel, Arle Richard, Burchardt, Aljoscha, Uszkoreit, Hans. Multidimensional Quality Metrics; A flexible System for Assessing Translation Quality, Berlin. 2013. URL: <http://www.mt-archive.info/10/Aslib-2013-Lommel.pdf> [10.9.2017.]
22. Makoushina, Julia. Translation Quality Assurance Tools: Current State and Future Approaches. 29.11.2007., Tomsk, Russia. URL: <http://www.mt-archive.info/Aslib-2007-Makoushina.pdf> [10.9.2017.]
23. Makoushina, Julia, Kockaert, J. Hendrick. Zen and the Art of Quality Assurance. 2008. URL: <http://www.mt-archive.info/Aslib-2008-Makoushina.pdf> [10.9.2017.]
24. Reinke, Uwe. State of the Art in Translation Memory Technology. Cologne University of Applied Sciences. 2013. URL: <http://www.blogs.uni-mainz.de/fb06-tc3/files/2015/11/25-145-3-PB.pdf> [5.9.2017.]
25. Ribeiro, A., Dias, G., Lopes, G., Mexia, J.: Cognates Alignment. U: Maegaard, B. (ed.) Proceedings of the Machine Translation Summit VIII (MT Summit VIII), Santiago de Compostela, Španjolska, European Association of Machine Translation, 2001. Str. 287–292. URL: <http://www.mt-archive.info/MTS-2001-Ribeiro.pdf> [10.9.2017.]
26. Sandrini, Peter. Localization and Translation. U: MuTra Journal. 2008. Innsbruck, Austrija. URL: <http://homepage.uibk.ac.at/~c61302/publik/localiz.pdf> [10.9.2017.]
27. Seljan, Sanja; Gašpar, Angelina; Pavuna, Damir. Sentence Alignment as the Basis for Translation Memory Database. U: 1. međunarodna znanstvena konferencija "The Future of Information Sciences: INFUTURE2007 – Digital Information and Heritage", 2007. Zagreb.
28. Simard, Michel; Plamondon, Pierre. Bilingual sentence alignment: balancing robustness and accuracy. Laval (Quebec), Canada. 1998.
29. Somers, Harold. Computers and Translation: A Translator's Guide. Philadelphia. John Benjamins Publishing Company, 2003.
30. Škrbina, Boris. Diplomski rad, Računalni alati u izradi prijevodne memorije, Zagreb, Filozofski fakultet, 2014. URL: <http://darhiv.ffzg.unizg.hr/id/eprint/4583> [10.9.2017.]
31. Yang, J.; Ciobanu, D.; Reiss, C; Secară, A. Using Computer Assisted Translation Tools' Translation Quality Assessment functionalities to assess students' translations. The Language

Scholar Journal. 2017. URL: <http://languagescholar.leeds.ac.uk/wp-content/uploads/sites/3/2017/05/Issue-1-YangReissSecaraCiobanu.pdf> [19.9.2017.]

32. Zerfass, Angelika. Evaluating Translation Memory Systems. 2002. <http://www.mt-archive.info/LREC-2002-Zerfass.pdf> [10.9.2017.]

Popis slika

Slika 1. Sravnjivanje u memoQ alatu	21
Slika 2. Greške u automatskom sravnjivanju.....	21
Slika 3. Spremanje segmenata u prijevodnu memoriju	22
Slika 4. Prijevodna memorija Poslovno dopisivanje en-hr TM.....	22
Slika 5. Ekstrakcija terminologije u memoQ alatu	23
Slika 6 Rezultat pokretanja ekstrakcije termina u memoQ-u	24
Slika 7. Terminološka baza Poslovno dopisivanje en-hr TB.....	25
Slika 8. Namještanje QA postavki - memoQ.....	29
Slika 9. QA u memoQ-u	29
Slika 10. Kreiranje projekta (Verifika)	31
Slika 11. Dodavanje terminološke baze u projekt	32
Slika 12. Pokretanje projekta (Verifika)	33
Slika 13. Rezultat provjere kvalitete (Verifika).....	33
Slika 14. Postavke provjere kvalitete u QAD-u.....	35
Slika 15. Pronađene greške u QAD-u	36
Slika 16. Ispravljanje grešaka u QAD-u	36
Slika 17. Pregledavanje pronađenih grešaka u QAD-u	37
Slika 18. Dodavanje datoteke u projekt	38
Slika 19. Dodavanje terminološke baze u projekt	39
Slika 20. Vrste provjera u alatu Xbench	40
Slika 21. Pronađene pravopisne greške	41
Slika 22. Rezultati provjere kvalitete u alatu Xbench.....	41
Slika 23. Postavke provjere kvalitete – memoQ.....	50
Slika 24. Uređivanje postavki za pisanje brojeva u hrvatskom jeziku – memoQ.....	51
Slika 25. Uređivanje postavki profila za hrvatski jezik – Verifika.....	51
Slika 26. Pogrešno pravilo pisanja raspona brojeva (Verifika)	52
Slika 27. Upravljanje listama provjere (Xbench)	53
Slika 28. Dodavanje novog pravila u listu provjere (Xbench).....	53
Slika 29. Uređivanje profila (QA Distiller)	54
Slika 30. Korištenje regularnih izraza u profilu (QA Distiller)	54
Slika 31. Prijedlog ispravka nađene greške (Verifika)	55
Slika 32. Pogrešan prijedlog ispravka (itnos u etnos, umjesto iznos) (memoQ).....	55
Slika 33. Jezgra MQM-a.....	57

Popis tablica

Tablica 1. Rezultati provjere kvalitete u memoQ alatu	30
Tablica 2. Rezultati provjere kvalitete u alatu Verifika	34
Tablica 3. Rezultati provjere kvalitete u alatu QA Distiller	37
Tablica 4. Rezultati provjere kvalitete u alatu Xbench	42
Tablica 5. Konačni rezultati QA alata.....	43
Tablica 6. Primjeri terminoloških grešaka	48
Tablica 7. Rezultati QA alata za evaluaciju kvalitete prijevoda	59