SVEUČILIŠTE U ZAGREBU
FILOZOFSKI FAKULTET
ODSJEK ZA INFORMACIJSKE I KOMUNIKACIJSKE ZNANOSTI
KATEDRA ZA DRUŠTVENO-HUMANISTIČKU INFORMATIKU

Analiza sustava za evaluaciju strojnih prijevoda primjenom Dinamičkog okvira i Višedimenzionalne metrike

diplomski rad

Mentor: dr. sc. Sanja Seljan, red. prof.
Ime studenta: Denis Kranjčić

Zagreb, 2016.

Sadržaj

Sažetak

Ovaj se diplomski rad bavi analizom sustava za evaluaciju strojnog prevođenja te se sastoji od teorijskog i praktičnog dijela. Na početku rada opisane su općenite primjene strojnog prevođenja i nekoliko konkretnih primjera. Nakon toga slijedi prikaz različitih vrsta evaluacije strojnih prijevoda, koje se u najširem smislu dijele na automatske i ljudske metode evaluacije. Potom su prikazane i opisane dvije platforme koje pružaju alate za ljudsku evaluaciju strojnih prijevoda: TAUS-ov dinamički okvir za evaluaciju kvalitete (engl. *TAUS Dynamic Quality Framework*) i Višedimenzionalne metrike za evaluaciju kvalitete (engl. *Multidimenskional Quality Metrics*). Detaljno su opisani alati koji su dio tih platformi te je objašnjeno postavljanje i obavljanje evaluatorskih zadataka. U praktičnom dijelu opisano je provedeno istraživanje u kojem se koristila većina opisanih alata. U istraživanju je provedena evaluacija strojnih prijevoda tekstova iz dvaju područja prevedenih pomoću alata Google Translate i Bing Translator. U zaključku rada analiziraju se rezultati istraživanja i funkcionalnosti korištenih alata. Na kraju rada nalazi se popis literature, popis slika, tablica i grafova te popis priloga.

ključne riječi: strojno prevođenje, evaluacija strojnog prevođenja, MQM, DQF, ljudska evaluacija, englesko-hrvatski jezični par

Abstract

This master's thesis deals with an analysis of systems for machine translation evaluation and consists of a theoretical and a practical section. The paper starts with a general description of the applications of machine translation and a few practical examples. This is followed by a presentation of different types of machine translation evaluation, which are most broadly divided into automatic and human methods of evaluation. After that, the paper focuses on two platforms which provide tools for human evaluation of MT: TAUS Dynamic Quality Framework and Multidimensional Quality Metrics. These tools are described in great detail, along with the process of setting up and performing evaluation tasks. For the practical part, an experimental research has been conducted in which most of the described tools were used. The research consists of an evaluation of machine translations of texts from two different domains performed by online tools Google Translate and Bing Translator. In the concluding section, the research results are analysed, as well as the functionalities of the tools used. At the end of the paper there is a list of works cited, a list of pictures, tables and graphs, and an appendix.

keywords: machine translation, machine translation evaluation, MQM, DQF, human evaluation, Croatian-English language pair

1. Uvod

Evaluacija strojnih prijevoda nije jednostavan zadatak. Kvaliteta prijevoda svojstveno je subjektivna pa je teško odrediti što je to objektivno i mjerljivo „dobar“ prijevod. Isti tekst moguće je prevesti na više dobrih načina, što znači da gotovo nikad ne postoji jedno, jedinstveno dobro rješenje. Bez obzira na to, postoji mnogo metoda evaluacije kojima se pokušavaju pomoću različitih metrika izmjeriti različiti parametri strojnih prijevoda.

U prvom dijelu ovog rada bit će ukratko opisane primjene strojnog prevođenja, koje uključuju primjene za diseminaciju, asimilaciju, razmjenu poruka u komunikaciji i sustave za pristup informacijama. Nakon toga bit će ukratko objašnjena automatska i ljudska evaluacija strojnih prijevoda. Za svaki od tih pristupa bit će ukratko opisan način rada, područja primjene te prednosti i nedostaci. Pozornost će se potom usmjeriti na dvije različite platforme za ljudsku evaluaciju: TAUS-ov Dinamički okvir (engl. *TAUS Dynamic Quality Framework*) i Višedimenzionalnu metriku za evaluaciju kvalitete (engl. *Multidimensional Quality Metrics*). U tom će se dijelu detaljno opisati način rada alata koje te platforme pružaju, strukture projektnih datoteka i rezultati koji se dobivaju na kraju evaluatorskih/anotatorskih zadataka. U zadnjem dijelu rada opisano je provedeno istraživanje, kojim se testira velik broj funkcionalnosti navedenih alata. U istraživanju se uspoređuju izlazi dvaju sustava za strojno prevođenje, odnosno alata Google Translate i Bing Translator, a korišteni su tekstovi iz dvaju različitih područja. Osim rezultata istraživanja u zaključku se također analizira učinkovitost i komplementarnost svih korištenih alata. Na kraju rada nalazi se popis literature, popis slika, tablica i grafova te popis priloga.

2. Primjene strojnog prevođenja

Strojno je prevođenje proces u kojem računalni program analizira tekst na jednom jeziku (izvorni tekst), i potom proizvodi tekst istoga značenja na drugom jeziku, bez učešća čovjeka u tome procesu.[1] Međutim, prevođenje kao takvo previše je složeno da bi sustavi za strojno prevođenje u potpunosti zamijenili ljudske prevoditelje. Takvi sustavi ni dan-danas u većini slučajeva nisu sposobni proizvesti izlaz koji se uopće može približiti kvaliteti ljudskog prijevoda. Osim toga, strojni prijevodi besplatnih *online alata* poput alata Google Translate često izazivaju podsmijeh zbog raznih značenjskih i gramatičkih pogrešaka koje se u njima mogu pojaviti.

No takav je pogled na strojno prevođenje pogrešan jer njegova svrha uopće nije zamjena čovjeka u prijevodnom procesu. Bez obzira na nedostatke općenitih sustava za strojno prevođenje koji su dostupni većini ljudi, slični sustavi imaju raznorazne primjene i mogu se koristiti u brojnim područjima.

2.1. Prijevodne potrebe

Strojno prevođenje u općenitom smislu može poslužiti za četiri različite namjene. Prva namjena uključuje korištenje strojnog prevođenja u svrhu diseminacije. To znači da se od strojnog prijevoda očekuje kvaliteta koja se inače očekuje od ljudskih prevoditelja, tj. da se ti prijevodi mogu objavljivati i prodavati ili interno distribuirati unutar neke tvrtke ili organizacije. Međutim, sustavi za strojno prevođenje proizvode izlaz koji u većini slučajeva trebaju naknadno pregledati i urediti ljudski prevoditelji ako se očekuje zadovoljavajuća razina kvalitete. Potreba za naknadnim uređivanjem često može biti znatna, što znači da sustav proizvodi samo početnu verziju prijevoda. Alternativni pristup može biti stroga propisanost forme ulaznog teksta, tj. korištene kontroliranog jezika, u kojem je definiran rječnik i struktura rečenica. U tom slučaju naknadno se uređivanje svodi na popravljanje manjeg broja sitnijih grešaka. Osim toga, neki su sustavi razvijeni za rad s vrlo usko definiranim vrstama tekstova te je za njih potrebno vrlo malo pripreme i naknadnog uređivanja teksta.[2]

[1] Što je to prijevod i prevođenje,a što je strojno prevođenje?. http://www.prevoditelj.com/prijevod-prevodjenje-i-strojno_prevodjenje/. 18.9.2016.
[2]Hutchins, J. The development and use of machine translation systems and computer-based translation tools. // International Symposium on Machine Translation and Computer Language Information Processing. Beijing: China, 1999. Str 26-28.

Druga je namjena strojnog prevođenja za svrhu asimilacije. Za tu je potrebu dostatna nešto manja razina kvalitete (pogotovo što se tiče stila) jer strojni prijevod u tom slučaju služi samo kako bi korisnik okvirno shvatio sadržaj određenog dokumenta, najčešće na najbrži mogući način. Budući da sustavi za strojno prevođenje najčešće ne mogu proizvesti prijevode visoke kvalitete, nekim je korisnicima dovoljna činjenica da mogu saznati informacije koje su im potrebne iz neuređenog izlaza. U nekim situacijama kada je prijevod loše kvalitete jedini način da se sazna značenje izvornika, strojno je prevođenje dostatan način za ispunjavanje te potrebe.[3]

Treća je funkcija strojnog prevođenja razmjena informacija između korisnika u komunikaciji jedan na jedan (npr. putem telefona ili u pisanoj korespondenciji). U današnje vrijeme potreba za ovakvim prijevodima sve više raste zbog velike količine tekstova na internetu, kao npr. na internetskim stranicama ili u elektroničkoj pošti. U takvom kontekstu potrebno je prenijeti osnovni sadržaj poruke, bez obzira na kvalitetu izlaza. S tom svrhom također se radi na razvoju sustava za prijevod govorenog teksta, poput onog u telefonskim razgovorima i poslovnim pregovorima. Problemi integracije prepoznavanja govora i automatskog prevođenja zasigurno su veliki, ali napredak u tom području svejedno postoji. U budućnosti se možda čak i mogu očekivati *online* sustavi za prijevod govora u vrlo ograničenim domenama.[4]

Četvrta namjena strojnog prevođenja ostvaruje se u sustavima za pristup informacijama. To uključuje integraciju prevoditeljskog softvera u sustavima za pretraživanje i dohvaćanje tekstualnih dokumenata iz baza podataka (npr. elektroničkih verzija članaka iz znanstvenih časopisa), sustavima za dohvaćanje bibliografskih podataka, sustavima za ekstrakciju informacija iz tekstova (npr. novinskih članaka), sustavima za sažimanje tekstova i sustavima za postavljanje upita u netekstualnim bazama podataka. Ovo je područje u fokusu brojnih europskih projekata, kojima je cilj širenje pristupa izvorima podataka i informacija za sve članove EU-a bez obzira na izvorni jezik.[5]

[3] Hutchins, J. The development and use of machine translation systems and computer-based translation tools. // International Symposium on Machine Translation and Computer Language Information Processing. Beijing: China, 1999. Str 26-28.
[4] ibid.
[5] ibid.

2.1. Korištenje strojnog prevođenja u organizacijama i institucijama

Usprkos njihovim ograničenjima, programi za strojno prevođenje koriste se u brojnim institucijama i organizacijama diljem svijeta. Jedan je od najvećih korisnika te tehnologije Europska komisija. Europska je unija, na primjer, uložila 2.375 milijuna eura u projekt projekta Sveučilišta u Göteborgu za izradu pouzdanog prevoditeljskog alata koji pokriva većinu jezika Europske unije.[6] Osim toga, Europska komisija u sklopu svog programa ISA izdvojila je 3,072 milijuna eura za izradu statističkog programa za strojno prevođenje MT@EC, koji je prilagođen administrativnim potrebama EU-a i koji bi trebao zamijeniti sustav za strojno prevođenje temeljen na pravilima koji se dosad koristio.[7]

Zbog prijetnje od terorizma u suvremenom svijetu, vojne snage u Sjedinjenim Američkim Državama također ulažu značajne količine novca u inženjering prirodnog jezika. Članovi vojne zajednice trenutačno su zainteresirani za prijevod i obradu jezika poput arapskog, paštunskog i darijskog. Vezano uz te jezike, fokus se stavlja na ključne fraze i brzu komunikaciju između članova vojske i civila putem aplikacija za mobilne telefone.[8] Osim toga, Američko zrakoplovstvo uložilo je preko milijun dolara za razvoj prijevodnih tehnologija.[9]

Mnoge tvrtke i institucije u svijetu koriste sustave za strojno prevođenje u kombinaciji s kontroliranim jezikom. Neke od poznatijih tvrtki koje se koriste takvim sustavima su Xerox Corporation, Caterpillar Corporation, Ford i General Electric.[10] Osim toga, neki sustavi koje prevode tekstove iz izuzetno uskih domena također su vrlo uspješni. Najpoznatiji je primjer takvog sustava kanadski sustav Météo, koji se koristi za prevođenje vremenskih prognoza s engleski na francuski i obrnuto. Sustavi s veoma ograničenim domenama pokazali su se uspješnima i u drugim industrijama, poput tekstilne, kemijske i elektroničke. [11]

[6] Multilingual Online Translation. http://www.molto-project.eu/. 19.9.2016.
[7] Machine Translation Service. http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm. 18.9.2016.
[8] Gallafent, A. Machine translation for the military. http://www.theworld.org/2011/04/machine-translation-military/. 18.9.2016.
[9] Jackson, W. Air Force wants to build a universal translator. https://gcn.com/articles/2003/09/09/air-force-wants-to-build-a-universal-translator.aspx. 18.9.2016.
[10] Hutchins, J. The development and use of machine translation systems and computer-based translation tools. // International Symposium on Machine Translation and Computer Language Information Processing. Beijing: China, 1999. Str 26-28.
[11] ibid.

3. Pristupi evaluaciji strojnih prijevoda

Razvoj sustava za strojno prevođenje najčešće je dugotrajan i težak proces za koji treba uložiti mnogo vremena i novca. Takvi se sustavi neprestano nadograđuju i proširuju, bez obzira radi li se o sustavima temeljenima na statistici ili na pravilima. Zbog neprestanog nadograđivanja i održavanja takvih sustava, često se provodi njihova evaluacija kako bi se vidjelo jesu li promjene utjecale na sustav pozitivno ili negativno. Evaluaciji strojnih prijevoda može se pristupiti na različite načine, a dvije su glavne kategorije evaluacije automatske metrike i ljudska evaluacija. Iako svaka od tih metoda ima svoje prednosti i nedostatke, bitno je napomenuti da se područja primjene tih metoda razlikuju, zbog čega je teško odrediti koja je od njih bolja. U ovom će se poglavlju prikazati i opisati te metode i njihove značajke.

3.1. Automatske metrike

Automatske metrike za evaluaciju rabe jedan ili više referentnih ljudskih prijevoda, koji se smatraju zlatnim standardom u kvaliteti prijevoda. Ti se referentni prijevodi koriste za uspoređivanje s izlazima sustava za strojno prevođenje ili prijevodnim kandidatima.[12] Postoji mnogo automatskih metoda za evaluaciju, od kojih su najpoznatije *Bilingual Evaluantion Understudy* (BLEU), *National Institute of Standards and Technology* (NIST), *Translation Error Rate* (TER), preciznost i odziv te *Metric for Evaluation of Translation with Explicit Ordering* (METEOR). Iako te metode računaju sličnost na različite načine, svima je zajedničko da boljom ocjenom ocjenjuju prijevodne kandidate koji su bliži ljudskom referentnom prijevodu.[13] Općenito se smatra da je automatska metoda bolja ako postiže visoku razinu korelacije s ljudskom procjenom.

Korelacija s ljudskom procjenom uglavnom se provjerava na dvije razine. Na razini rečenice metrika računa ocjenu za prevedenu rečenicu, koja se tada korelira s ljudskom procjenom za istu rečenicu. Na korpusnoj razini računa se ukupna ocjena svih rečenica za ljudsku procjenu i

[12] Second Machine Translation Marathon. Bilješke s predavanja. Njemačka : Berlin, 2008. citirano u Brkić, Marija; Vičić, Tomislav; Seljan, Sanja. Evaluation of the Statistical Machine Translation Service for Croatian-English. // 2nd international conference: The future of information sciences (INFuture 2009) : Digital resources and knowledge sharing : proceedings / Stančić, Hrvoje ; Seljan, Sanja ; Bawden, David ; Lasić-Lazić, Jadranka ; Slavić, Aida (ur.). Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009. 319-332

[13] Jurafsky, D.; Martin H. J. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. New Jersey: Pearson education, 2009.

procjenu metrike te se tada te ukupne ocjene koreliraju. Rezultati za korelaciju na razini rečenice rijetko se objavljuju, iako npr. Banjerjee i Lavie (2005)[14] prikazuju rezultate korelacije koji pokazuju da je, za njihovu metriku, korelacija na razini rečenice znatno lošija od korelacije na korpusnoj razini.

Banjerjee i Lavie (2005)[15] navode pet svojstava koja svaka dobra automatska metrika mora posjedovati: korelaciju, osjetljivost, dosljednost, pouzdanost i općenitost. Dobra korelacija znači da metrika mora davati rezultate koji su što bliži ljudskoj procjeni, a dosljednost da daje slične rezultate za isti sustav i sličan tekst. Nadalje, metrika mora biti osjetljiva na razlike između sustava tako da su oni sustavi koji su slično ocjenjeni također oni koji rade jednako dobro. Općenitost metrike odnosi se na njezinu sposobnost da se može koristiti s tekstovima iz različitih područja, u širokom rasponu situacija i zadataka.

Automatska metrika najčešće se koristi u razvoju sustava za strojno prevođenje kada je potrebno na brz i jeftin način utvrditi kako promjene u sustavu utječu na izlaz sustava. Međutim, ti rezultati nisu nužno informativni za ljudske korisnike jer ne govore ništa o tome u čemu je konkretno problem s prijevodom.[16] Računanje korelacije s nekim referentnim prijevodom također nije uvijek najbolji pristup evaluaciji zbog toga što za određeni segment najčešće postoji više od jednog dobrog prijevoda. Nadalje, automatska metrika koja dobro radi s jednom vrstom teksta nije uvijek primjenjiva na neku drugu vrstu, koja se može razlikovati po stilu ili razini slobode izražaja. Zbog takvih nedostataka automatskih metrika, za određene se primjene i svrhe istraživanja koriste metode ljudske evaluacije.

[14] Banerjee, S.; Lavie, A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements. Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Comupational Linguistics. Michigan, 2005.
[15] ibid.
[16] Lommel, A. Multidimensional Quality Metrics (MQM): A New Framework for Translation Quality Assessment. Prezentacija. 2014.

3.2. Ljudska evaluacija

Ljudska evaluacija strojnih prijevoda dolazi u nekoliko različitih oblika, od kojih su najčešći ocjenjivanje kvalitete, ocjenjivanje adekvatnosti, ocjenjivanje fluentnosti, rangiranje, i analiza pogrešaka.[17] Kod ocjenjivanja kvalitete, evaluatori pridružuju ocjene prijevodima na temelju predodređene skale vrijednosti. Na primjer, može se koristiti skala od 1 do 5, u kojoj je 1 najniža ocjena, a 5 najviša. Jedan je od izazova ovog pristupa utvrđivanje jasnog opisa svake vrijednosti na skali i preciznih razlika između razina kvalitete. Čak i kada postoje eksplicitne smjernice za evaluaciju, evaluatori mogu imati problema s pridruživanjem numeričkih vrijednosti kvaliteti prijevoda.[18]

Ocjenjivanje adekvatnosti također se svodi na pridruživanje brojčanih ocjena prijevodu. U ovom slučaju evaluatori ne ocjenjuju kvalitetu, pojam koji je općenito teško definirati, već adekvatnost, koju *Linguistic Data Consortium*[19] definira kao „količinu značenja izraženu u referentnom prijevodu ili izvornom tekstu koja je također izražena u ciljnom tekstu“. Za ovakav evaluatorski zadatak evaluatori trebaju izvrsno poznavati izvorni i ciljni jezik kako bi mogli procijeniti jesu li informacije dobro sačuvane u prijevodu.

Ocjenjivanje fluentnosti uključuje samo ciljni tekst, bez gledanja izvornika. Kriteriji ocjenjivanja su gramatika, pravopis, izbor riječi i stil, a najčešće se koristi skala od jedan do pet, gdje ocjena 5 označava da je tekst besprijekoran, a ocjena 1 da je nerazumljiv.

Rangiranje je zadatak u kojem su evaluatorima predstavljena dva ili više izlaza različitih sustava za strojno prevođenje, a oni moraju izabrati najbolji ili ih rangirati ovisno o kvaliteti. Takvi zadaci mogu biti zbunjujući za evaluatore u situacijama kad su segmenti koji se rangiraju gotovo identični ili kad sadrže različite tipove pogrešaka koje je teško uspoređivati. U takvim slučajevima, evaluatori moraju odlučiti koje pogreške imaju veći utjecaj na kvalitetu prijevoda.[20] Međutim,

[17] Pospelova, O.; Rowda, J. Human Evaluation of Machine Translation. 26.6.2016. http://www.ebaytechblog.com/2016/06/26/human-evaluation-of-machine-translation/. 6.9.2016.
[18] Koehn, P.; Monz, C. Manual and Automatic Evaluation of Machine Translation between European Languages. // Proceedings of the Workshop on Statistical Machine Translation. New York City : Association for Computational Linguistics, 2006. Str. 102-121.
[19] Adequacy/Fluency Guidelines. Svibanj 2013. *Quality Evaluation using Adequacy and/or Fluency Approaches*. https://www.taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluency-guidelines.6.9.2016.
[20] Denkowski, M.; Lavie, A. Choosing the Right Evaluation for Machine Translation: an Examination of Annotator and Automatic Metric Performance on Human Judgement Tasks. // Proceedings of the Ninth Biennial Conference of the Association for Machine Translation in the Americas. Denver, Colorado. 2010.

utvrđeno je da je ljudima lakše rangirati sustave nego ih ocjenjivati[21], a razlog tome su već spomenuti problemi s pokušajima kvantifikacije kvalitete prijevoda.

Analiza pogrešaka znatno je kompleksniji zadatak od prethodno navedenih. Evaluatori moraju prepoznati i klasificirati pogreške u izlazu sustava za strojno prevođenje, kao npr. izostavljene riječi, pogrešan red riječi, slaganje u rodu, broju i padežu itd. Klasifikacijska shema pogrešaka uglavnom ovisi o jeziku, vrsti sadržaja i cilju istraživanja. Problem je s ovakvom vrstom evaluacije u tome što je mogu obavljati samo stručni lingvisti, dok za ostale metode ljudske evaluacije to nije potrebno.

Osim već navedenih problema, ljudska evaluacija ima i nekih drugih nedostataka. Kao prvo, takva je evaluacija skup i dugotrajan proces, pogotovo u usporedbi s automatskom evaluacijom. Sam postupak evaluacije može potrajati satima (ovisno o količini podataka i vrsti zadatka), ali mnogo vremena također odlazi na pripremu, edukaciju evaluatora, sastavljanje smjernica itd. Kako bi rezultati bili što reprezentativniji, preporuča se da isti evaluatorski zadatak radi što više evaluatora. Njih najčešće treba platiti, a cijena evaluacije eksponencijalno raste što je veći broj ljudi koji sudjeluju na zadatku. Problem također postoji s razinom dosljednosti u označavanju jednog evaluatora (*intra-rater agreement*) i razinom slaganja između različitih evaluatora (*inter-rater agreement*).

Bez obzira na sve ove nedostatke, ljudska evaluacija izrazito je koristan način za utvrđivanje kvalitete prijevoda. Budući da kompetentni ljudski evaluatori mogu u pravilu mnogo bolje odrediti kvalitetu i probleme u određenom tekstu, ova je metoda pouzdanija od korištenja automatskih metrika. Analiza pogrešaka u ljudskoj evaluaciji najbolji je primjer razine evaluacije koju automatske metrike ne mogu postići. Osim toga, za ljudsku evaluaciju najčešće nisu potrebni referentni prijevodi i paralelni tekstovi. To joj daje veliku prednost pred automatskom evaluacijom zbog toga što takvi resursi nisu uvijek dostupni.

U nastavku rada bit će prikazane dvije novije platforme za ljudsku evaluaciju prijevoda. Prva je platforma TAUS-ov Dinamički okvir za evaluaciju kvalitete, koji nudi mnoštvo resursa i alata vezanih uz prijevodnu industriju. Budući da je svrha ovog rada prikazivanje alata za evaluaciju strojnih prijevoda, najveća pozornost bit će posvećena alatima koji dolaze u sklopu

[21] Vilar, D.; Leusch, H. N.; Banchs, R. Human evaluation of machine translation through binary system comparisons. // *ACL2007 SMT Workshop*. 2007.

Dinamičkog okvira i njihovim funkcionalnostima. Osim opisa alata i njihovih funkcionalnosti, također će se opisati cjelokupni proces postavljanja zadataka i njihova izvršavanja.

Druga platforma koja će biti opisana u radu je Višedimenzionalna metrika (MQM), koja je nastala kao dio projekta QTLaunchPad. Radi boljeg razumijevanja ovog projekta, najprije će se opisati LISA-in model za osiguranje kvalitete, koji je poslužio kao temelj za stvaranje tih metrika. Nakon toga će se opisati vrste problema koji mogu biti sadržani u Višedimenzionalnoj metrici i na koji se način ti problemi mogu pridruživati problemima u tekstu. Višedimenzionalna metrika može se koristiti u nekoliko različitih alata, pa će također biti prikazan način rada tih alata, a uz to i struktura projektnih datoteka koje su potrebne za postavljanje anotatorskog zadatka. Osim toga, bit će prikazana i objašnjena formula prema kojoj se računaju ocjene tekstova označenih pomoću Višedimenzionalne metrike.

4. TAUS-ov Dinamički okvir za evaluaciju kvalitete

TAUS je resursni centar za globalnu jezičnu i prijevodnu industriju, a njegova je misija poboljšavanje prevođenja kroz inovacije i automatizaciju.[22] U lipnju 2012. TAUS je pokrenuo svoj Dinamički okvir za evaluaciju kvalitete (*Dynamic Quality Framework,* DQF), koji omogućuje postavljanje standarda tako što pruža okvir prikladnih modela za evaluaciju kvalitete na temelju vrste sadržaja, namjene korištenja, alata, procesa i drugih varijabli. Dinamički okvir zapravo je baza znanja koja dokumentira najbolje prakse u primjeni evaluacijskih modela i zajedničkih alata u prevoditeljskoj industriji.[23] Osim najboljih praksi, DQF također sadrži izvješća, predloške i nekoliko alata za evaluaciju prijevoda, primjenjivih na strojne i ljudske prijevode. Ti alati omogućuju prevoditeljima da uspoređuju prijevode, da procjenjuju njihovu točnost i fluentnost, da mjere produktivnost naknadnog uređivanja (*post-editing*) te da ocjenjuju prevedene segmente na temelju tipologije pogrešaka.

U studenom 2014. DQF je postao dio TAUS-ove Platforme za evaluaciju (*TAUS Evaluate platform*). Ova platforma sadrži sveobuhvatan skup alata, demo verzija, primjera korištenja, metrika, izvješća i podataka koji pomažu postavljanju standarda u industriji. DQF-ov čarobnjak za profiliranje sadržaja (*DQF Content profiling wizard*) koristi se za pomoć pri izboru modela za evaluaciju kvalitete najprikladnijeg za specifične zahtjeve. Čarobnjak se koristi već spomenutom bazom znanja, koja je u obliku wikija (vrsta sustava za upravljanje sadržajem), što znači da korisnici mogu otvoreno uređivati njezin sadržaj. Korisnici također mogu raspravljati o prevoditeljstvu učlanjivanjem u Zajednicu za evaluaciju (*Evaluate Community*) i sudjelovati u *webinarima* o kvaliteti prijevoda.

[22] TAUS - Mission. https://www.taus.net/mission. 6.9.2016.
[23] TAUS launches Dynamic Quality Evaluation Framework. 12.6.2016. https://www.taus.net/think-tank/news/press-release/taus-launches-dynamic-quality-evaluation-framework. 6.9.2016.

4.1. DQF alati[24]

Kako je već spomenuto, TAUS-ov Dinamički okvir sadrži nekoliko alata za evaluaciju strojnog i ljudskog prevođenja. Tim se alatima može pristupiti putem kontrolne ploče za kvalitetu (*Quality Dashboard*) nakon registracije. Alatima se ne pristupa direktno, već se za svaki zadatak stvara novi projekt, kojemu je potrebno navesti ime, vrstu, ime tvrtke, vrstu sadržaja, industriju te izvorni i ciljni jezik. Ponuđene vrste projekata su rangiranje i uspoređivanje, testiranje produktivnosti te evaluacija kvalitete. Svaka od ovih kategorija dijeli se na određene podgrupe, koje će biti opisane u nastavku.

4.1.1. Rangiranje i uspoređivanje

Rangiranje i uspoređivanje je zadatak koji omogućuje korisnicima da izaberu sustav za strojno prevođenje ili ljudskog prevoditelja na temelju kvalitete izlaza. Drugim riječima, evaluatorima predstavlja izlaze dvaju ili triju sustava za strojno prevođenje segment po segment te od njih traži da odluče koji je najkvalitetniji. Postoje dva različita načina rada, a to su brzo uspoređivanje (*Quick Comparison*) i uspoređivanje rangiranjem (*Rank Comparison*). Kod brzog uspoređivanja, evaluator odabire samo jedan izlaz koji smatra najkvalitetnijim, dok u uspoređivanju rangiranjem svakom segmentu pridodaje ocjenu od 1 (najbolji prijevod) do 3 (najgori prijevod). Izlazi koje evaluator ocjenjuje poredani su nasumično i evaluator ne zna koji je segment izlaz kojeg sustava. Bitno je naglasiti logičan zaključak da u slučaju kada se evaluiraju samo dva sustava, nema smisla koristiti uspoređivanje rangiranjem, iako je to tehnički moguće.

Nakon što odabere vrstu zadatka, osoba koja postavlja evaluacijski zadatak treba priložiti projektnu datoteku u kojoj se nalaze segmenti za evaluaciju. Ta datoteka mora biti u xls, xlsx ili csv formatu, a njezinu strukturu trebaju sačinjavati sljedeći stupci:

- identifikacijski broj (*ID*): sadrži identifikator segmenta. Tijekom evaluacije segmenti se prikazuju po određenom redu, ovisno o identifikacijskom broju
- izvorni segment (*Source Segment*): sadrži rečenicu izvornika
- podrijetlo segmenta (*Segment Origin*): sadrži ime dokumenta iz kojeg rečenica potječe. To je ime prikazano u polju *Filename* (ime datoteke) koje se može vidjeti tijekom evaluacije.

[24] Opisi alata dolaze iz autorova iskustva s alatima i uputama koje su dostupne pri postavljanju zadataka. Budući da se upute otvaraju u istom prozoru u kojem se namještaju postavke zadataka, nije bilo moguće citirati konkretnu URL adresu.

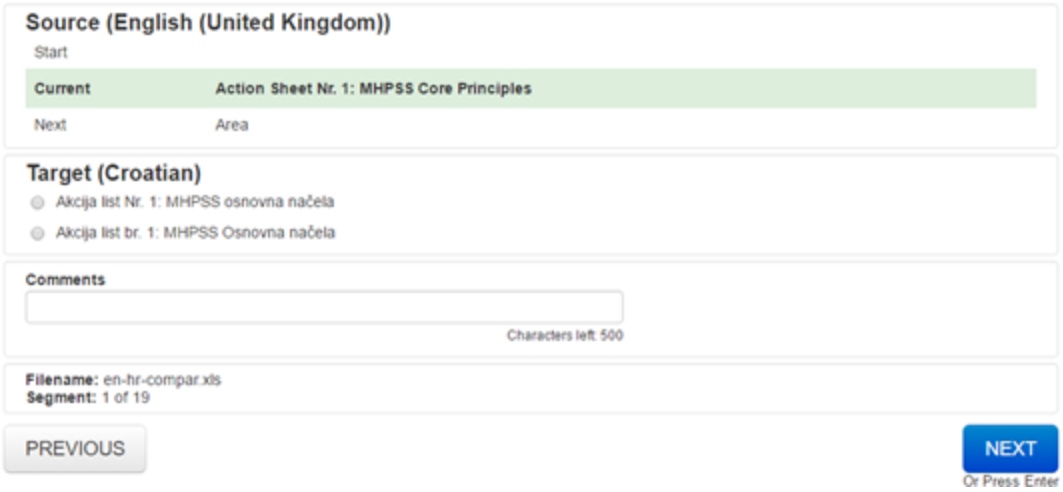
- izlaz prvog sustava (*Engine 1 Output*): sadrži rečenicu na ciljnom jeziku koju će evaluatori uspoređivati s ostalim rečenicama na ciljnom jeziku. Najčešće se u ovom stupcu nalazi nepromijenjeni izlaz sustava za strojno prevođenje koji treba usporediti s izlazima ostalih sustava
- izlaz drugog sustava (*Engine 2 Output*)
- izlaz trećeg sustava (*Engine 3 Output*).

Maksimalni broj sustava koje je moguće ocjenjivati je tri, a bitno je i napomenuti da se broj sustava koji se uspoređuju ne navodi na posebnom mjestu nego ovisi isključivo o tome koliko je sustava navedeno u projektnoj datoteci. Ako je datoteka u csv formatu, njezina struktura je ista, ali potrebno je paziti da tekst bude kodiran kao UTF-8 i da su stupci odijeljeni tabulatorom. Na sljedećoj je slici prikazano kako struktura projektne datoteke izgleda u praksi.

	A	B	C	D	E
1	ID	Source Segment	Segment Origin	Google Translate	Bing Translator
2	2	Action Sheet Nr. 1: MHPSS Core Principles	OPSIC_MHPSS.docx	Akcija list br. 1: MHPSS Osnovna načela	Akcija list Nr. 1: MHPSS osnovna načela
3	3	Area	OPSIC_MHPSS.docx	područje	Područje
4	4	All event types, all target groups, all phases	OPSIC_MHPSS.docx	Svi tipovi događaja, sve ciljne skupine, u svim fazama	Svih vrsta događaja, sve ciljane skupine, sve faze
5	5	MHPSS core principles in both IASC and NATO TENTS guidelines1,2	OPSIC_MHPSS.docx	temeljna načela MHPSS u oba IASC-a i NATO šatora guidelines1,2	MHPSS osnovnih načela u IASC i NATO-a šatora guidelines1, 2
6	6	Principle 1: Ensure human rights and equity	OPSIC_MHPSS.docx	1. načelo: Osigurati ljudska prava i jednakost	Princip 1: Osigurati ljudska prava i jednakost

Slika 1. Struktura projektne datoteke – rangiranje i usporedba (TAUS)

Nakon što priloži projektnu datoteku, upravitelj projekta šalje zadatak evaluatorima putem elektroničke pošte. Evaluatori primaju poveznicu na evaluatorski zadatak i upute o tome kako ga treba izvršiti. Ovaj je dio procesa identičan za sve vrste evaluatorskih zadataka u sklopu Dinamičkog okvira. U samom evaluatorskom zadatku, evaluatori uspoređuju ili rangiraju prijevode ekvivalentnih segmenata iz izvornog jezika sve dok ne završe sa svim segmentima koji su navedeni u projektnoj datoteci. Primjer sučelja brze usporedbe može se vidjeti na sljedećoj slici.



Slika 3: DQF - Brzo uspoređivanje

Kada evaluator završi sa zadatkom, upravitelj zadatka putem elektroničke pošte prima obavijest u kojoj se nalaze dvije poveznice. Prva poveznica vodi na stranicu na kojoj se mogu pregledati podaci o projektu i status riješenosti zadatka, koji pokazuje kako evaluatori napreduju sa zadatkom. Druga poveznica omogućava pristup izvješću o projektu, gdje se mogu pregledati razni statistički podaci u obliku grafikona koji prikazuju koji je sustav označen kao bolji u odnosu na druge. Ovi se podaci mogu pregledavati pojedinačno za svakog evaluatora ili kao prosjek rezultata svih evaluatora.

4.1..2. Testiranje produktivnosti

Druga je vrsta projekta/evaluacije koju omogućavaju DQF alati testiranje produktivnosti. Ovo testiranje omogućuje usporedbu razlike u brzini između naknadnog uređivanja izlaza sustava za strojno prevođenje i ljudskog prevođenja od nule. Međutim, ugađanjem različitih parametara moguće je također mjeriti neke druge vrijednosti, poput brzine ljudskog prevođenja (korisniku su ponuđeni samo segmenti iz izvornog jezika, koje treba prevesti), brzine samo naknadnog uređivanja strojnog prijevoda, brzine ljudskog prevođenja uz pomoć prijevodne memorije te brzine ljudskog prevođenja i naknadnog uređivanja uz pomoć strojnog prevođenja i prijevodne memorije. Osim toga, moguće je i zadati kakva treba biti kvaliteta naknadnog uređivanja strojnog prijevoda (dovoljno dobra ili slična/jednaka ljudskom prijevodu), što se u evaluatorskom zadatku manifestira kao uputa za evaluatora, ali nema velikog utjecaja na izgled i strukturu zadatka.

Projektna datoteka koju upravitelj zadatka mora postaviti za testiranje produktivnosti po strukturi se malo razlikuje od one za uspoređivanje i rangiranje, iako su podržani formati datoteke isti. Budući da se ovdje ne uspoređuje više outputa, nego strojni prijevod služi za naknadno uređivanje, u datotekama se navodi izlaz samo jednog sustava za strojno prevođenje. Ostali stupci jednaki su stupcima koji su navedeni u opisu prethodnog zadatka. Bitno je također napomenuti da je ovakva struktura podataka u datoteci jednaka i za treći alat u dinamičkom okviru, evaluaciju kvalitete.

Nakon završetka zadatka, voditelj projekta može pristupiti izvješću koji pokazuje nekoliko grafova ovisno u postavkama zadatka. Ovdje se, na primjer, može pregledati usporedba produktivnosti naknadnog uređivanja i prevođenja te usporedba produktivnosti različitih prevoditelja, što je oboje izraženo u broju riječi po satu. Također se može vidjeti prosjek vremena potrebnog za prijevod/naknadno uređivanje u odnosu na dužinu rečenice, broj riječi po satu u odnosu na dužinu rečenice i graf udaljenosti uređivanja (*edit distance graph*), koji prikazuje koliko je truda bilo potrebno za dovođenje ponuđenog prijevoda (prijevodne memorije ili strojnog prijevoda) na željenu razinu kvalitete (uloženi trud računa se prema Levenshteinovu algoritmu normaliziranom ovisno o duljini segmenta). Osim grafova, moguće je preuzeti i tabličnu datoteku u kojoj se za svaki segment vide prevedeni ili naknadno uređeni segmenti, dužina segmenta, vrijeme naknadnog uređivanja i uložen trud.

4.1.3.. Evaluacija kvalitete

U TAUS-ovu Dinamičkom okviru evaluacija kvalitete zapravo je krovni termin za tri različite vrste zadataka: evaluaciju fluentnosti, evaluaciju adekvatnosti i označavanje tipoloških pogrešaka. Podvrsta zadatka odabire se pomoću potvrdnog okvira (*checkbox*), a moguće je zadatak postaviti na način da uključuje više podzadataka odjednom; ako se odaberu sva tri, evaluator treba za svaki segment ocjenjivati fluentnost i adekvatnost te utvrđivati tipološke pogreške.

Evaluacija *fluentnosti* omogućava ocjenjivanje svakog segmenta ocjenom od 1 do 4 ovisno o tome koliko je prijevod gramatički i pravopisno točan te koliko intuitivno i prirodno zvuči izvornom govorniku ciljnog jezika. Pritom su ocjene definirane na sljedeći način:

4	Besprijekorno	Tekst je tečan i bez pogrešaka.
3	Dobro	Tekst dobro teče, ali sadrži nekoliko sitnih pogrešaka.
2	Nefluentno	Tekst je loše napisan i teško ga je razumjeti.
1	Nerazumljivo	Tekst je izuzetno loše napisan i nemoguće ga je razumjeti.

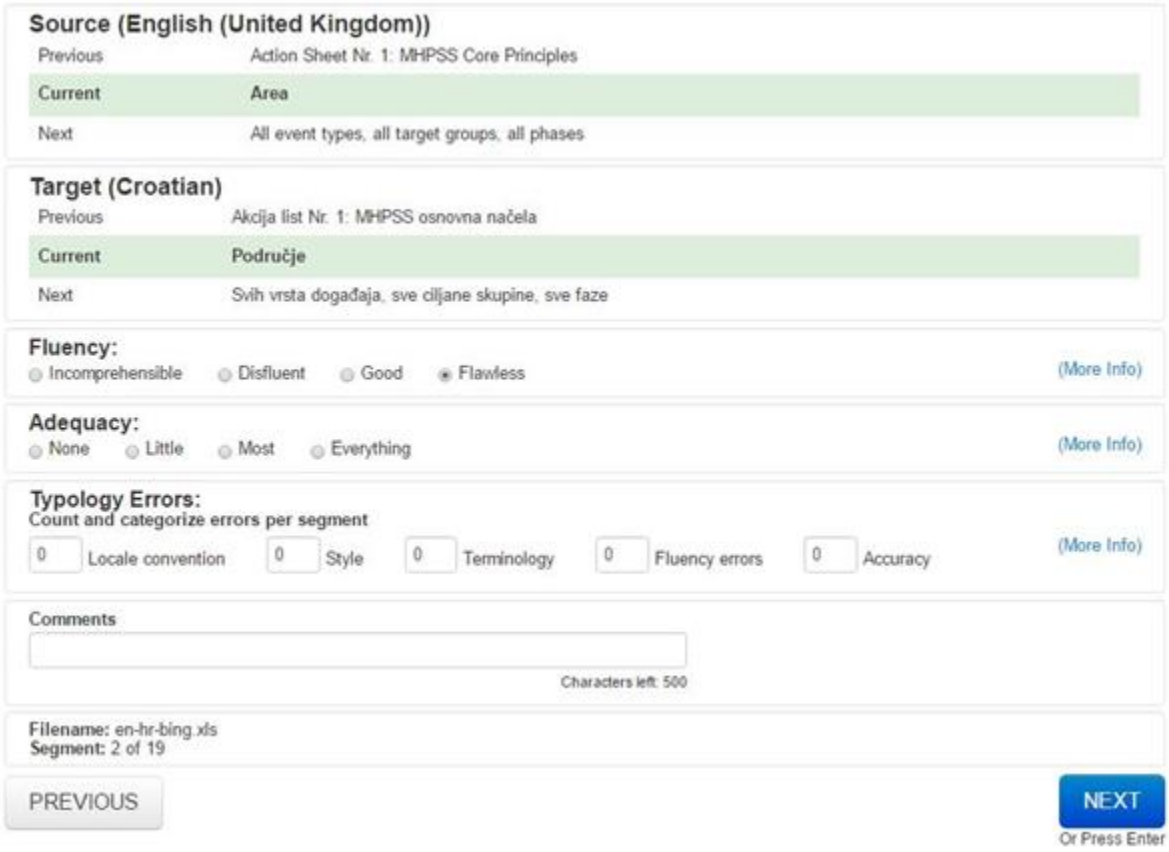
Tablica 1: Fluentnost - raspon ocjena

Evaluacija *adekvatnosti* također omogućava ocjenjivanje segmenata ocjenom od 1 do 4, no u ovom se slučaju ocjenjuje koliko je značenje izvornog teksta dobro preneseno u ciljni jezik. Ocjena 4 označava da je značenje u potpunosti preneseno, ocjena 3 da je većina značenja prenesena, ocjena 2 da se u ciljnom segmentu nalaze samo djelići značenja izvornika, a ocjena 1 da značenje izvornika nije nimalo preneseno u ciljni jezik.

Zadnja podvrsta zadatka, *utvrđivanje tipoloških pogrešaka*, omogućava kategoriziranje i brojanje prijevodnih pogrešaka na razini segmenta. Tipologija pogrešaka preuzeta je iz Višedimenzionalnih metrika, no prisutna su samo četiri tipa pogrešaka (problema): točnost, fluentnost, terminologija, stil i konvencije regionalne sheme (*locale conventions*)[25]. Granularnije vrste pogrešaka spominju se samo u uputama kao pomoć za lakše određivanje koji tip pogreške podrazumijeva koje podvrste pogrešaka. Za razliku od anotacije pomoću translate5 alata, gdje se pogreške označavaju izravno na riječi ili dijelu segmenta gdje se pojavljuju, ovdje se pogreške samo prebrojavaju i upisuje se njihov broj za svaki segment.

[25] konvencije regionalne sheme odnose se na načine pisanja brojeva, datuma, valuta itd. u ciljnoj kulturi

Na sljedećoj je slici prikazan prozor zadatka evaluacije kvalitete u kojem se provjeravaju fluentnost, adekvatnost i tipologija pogrešaka. Već navedene informacije o vrstama zadataka i značenja ocjena također se mogu u svakom trenutku pregledavati izravno iz tog prozora.



Slika 4: Evalucija kvalitete

Kad evaluator završi zadatak, upravitelj projekta prima posebno izvješće za svaku podvrstu zadatka koju je uključio u postavkama. U izvješću za test fluentnosti nalazi se grafikon koji prikazuje ukupnu distribuciju ocjena ovisno o broju segmenata i grafikoni koji prikazuju distribuciju ocjena za svakog evaluatora posebno. Izvješće za test fluentnosti izgleda slično, dok se u izvješću za tipološke pogreške nalazi tablica s ukupnim zbrojem pogrešaka u zadatku za svaku vrstu pogreške.

5. Višedimenzionalna metrika za evaluaciju kvalitete

Sustav Višedimenzionalnih metrika za evaluaciju kvalitete (*Multidimensional Quality Metrics,* MQM) razvijen je kao dio projekta QTLaunchPad, koji financira Europska Unija. Cilj je tog projekta nadogradnja modela za osiguravanje kvalitete kakav je u prošlosti razvijala udruga za standarde u lokalizaciji LISA, koja je prestala s radom 28. veljače 2011.[26] Specifikacija LISA-inog modela sastojala se od popisa vrsta pogrešaka koje je bilo moguće povezati sa specifičnim pogreškama u tekstovima. Različite vrste pogrešaka u ovom modelu imale su različitu težinu, npr. pravopisna pogreška smatrala se manje problematičnom od značenjske itd. Pogreške bi se na kraju označavanja zbrajale sa svrhom pružanja ocjene kvalitete (uglavnom prikazane u obliku postotka), što je pružalo klijentima i pružateljima usluga mogućnost da postave prag kvalitete i na taj način provjere zadovoljavaju li prijevodi njihove potrebe.[27]

Međutim, LISA-in model za osiguranje kvalitete pružao je jedinstven, nefleksibilan skup metrika koje nisu odgovarale svim raznovrsnim potrebama korisnika. Zbog toga su korisnici trebali prilagođavati specifikacije svojim potrebama i nisu ih se strogo pridržavali, što znači da evaluacije različitih korisnika nije bilo moguće uspoređivati. To je također značilo da su se prakse u evaluaciji razlikovale od korisnika do korisnika na gotovo nasumičan način. Osim toga, u istraživanju koje je provedeno u sklopu projekta QTLaunchPad, ispostavilo se da su metrike koje su razvijene za određene projekte u organizacijama i koje su kasnije te organizacije usvojile za druge projekte često korištene usprkos tome što su bile problematične i neprimjenjive u novim situacijama.[28]

Sustav Višedimenzionalnih metrika razvijen je s ciljem da popravi nedostatke prethodnih načina evaluacije kvalitete. Ovaj sustav vuče korijene iz pokušaja nadogradnje LISA-inog modela za osiguravanje kvalitete, što znači da preuzima velik broj svojih načela iz tog modela. No za razliku od prethodnih metrika za evaluaciju kvalitete, MQM je od početka dizajniran da bude

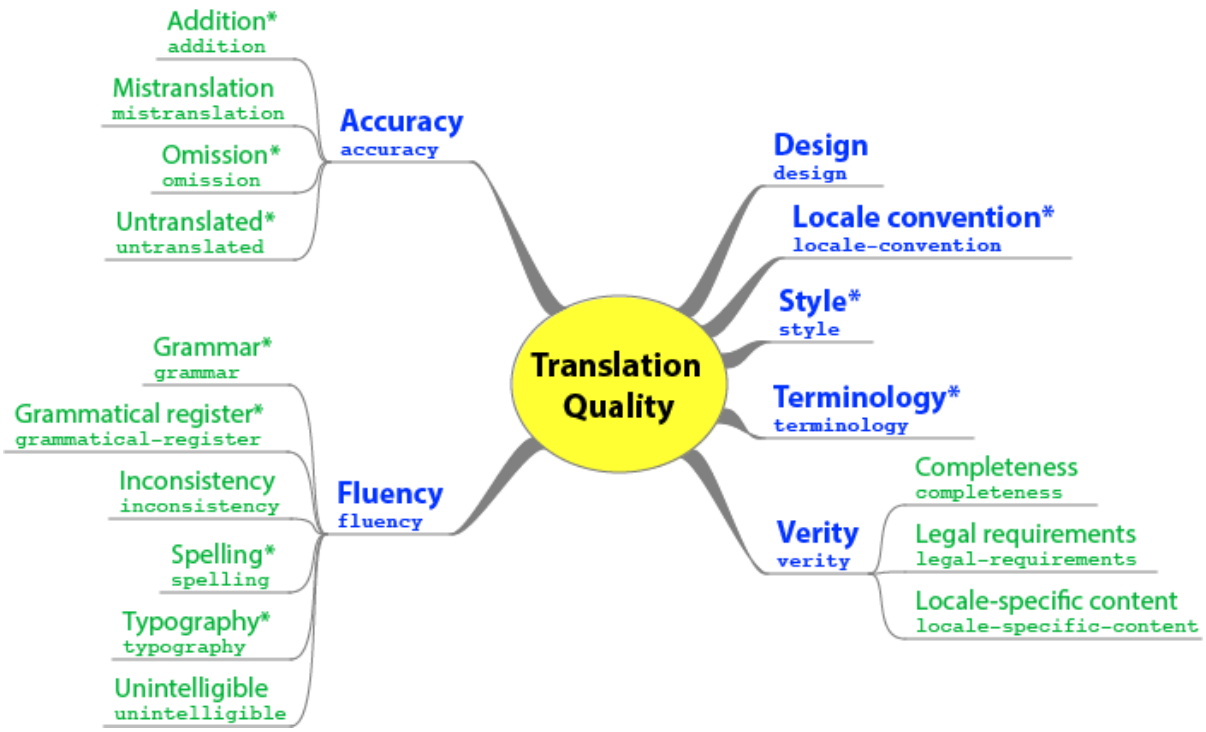
[26] DePalma, D. A. LISA Shuts Down Operations. 28. 2. 2011. http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=1357&moduleId=390. 6.9.2016.
[27] Lommel, A.; Uszkoreit, H.; Burchardt, A. Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. Revista Tradumàtica: tecnologies de la traducció. 2014. Str. 455-462.
[28] ibid.

fleksibilan i poveziv s drugim standardima kako bi se evaluacija kvalitete integrirala u cjelokupni ciklus nastanka dokumenta.[29]

Središnja je komponenta MQM-a hijerarhijski popis vrsta problema, koji je rezultat proučavanja postojećih metrika za evaluaciju kvalitete i problema koji se nalaze u alatima za automatsku provjeru kvalitete. Problemi su ograničeni na samo one koji imaju veze s jezikom i formatom, što znači da su izostavljeni problemi povezani s kvalitetom projekta (npr. poštivanje rokova ili dovršenost projekta), koji se nalaze u LISA-inom modelu za procjenu kvalitete.[30] Trenutna verzija MQM hijerarhije sadrži preko 100 vrsta problema[31] koji predstavljaju generalizirani nadskup problema koji se nalaze u postojećim metrikama i alatima. Međutim, iako su neki problemi generalizirani, moguće je implementirati potprobleme prema potrebi kao prilagođene ekstenzije za MQM.

Na najvišoj razini hijerarhije, MQM je podijeljen na 10 vrsta problema: točnost (*accuracy*), dizajn (*design*), fluentnost (*fluency*), konvencije regionalne sheme (*locale convention*), stil (*style*), terminologija (*terminology*), istinitost (*verity*), kompatibilnost (*compatiblity*) i ostalo. Ove vrste problema dijele se na brojne podvrste pa je puna hijerarhijska shema MQM-a izrazito velika. Međutim, bitna značajka MQM-a je to što se ni u jednom projektu ne koriste sve vrste problema, nego samo one koje su potrebne i informativne za određeni projekt. Kako bi se pojednostavila primjena MQM-a, definirana je manja „jezgra“ (*MQM Core*), koja se sastoji od 20 najčešćih vrsta problema koji se pojavljuju u analizi kvalitete prevedenih tekstova. Ta jezgra seže do relativno visoke razine zrnatosti koja je prikladna za većinu zadataka, a korisnike MQM-a se potiče da koriste probleme iz jezgre kad god je to moguće kako bi se omogućila što veća interoperabilnost među sustavima.[32] Iako čak i sama jezgra sadrži više problema nego što je vjerojatno da će biti potrebno za bilo koju konkretnu primjenu, korisnici mogu definirati podskupove jezgre sukladno njihovim potrebama. Također je moguće definirati i manju razinu zrnatosti ili izbaciti neke od vrsta problema ovisno o potrebi, ali preporuča se da svaki zadatak sadrži barem točnost i fluentnost. Na sljedećoj se slici nalazi grafički prikaz jezgre MQM-a.

[29] Ibid.
[30] Ibid.
[31] Multidimensional Quality Metrics (MQM) Issue Types. http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html. 6.9.2016.
[32] Multidimensional Quality Metrics (MQM) Definition. http://www.qt21.eu/mqm-definition/definition-2015-12-30.html. 6.9.2016.



Slika 5: Jezgra MQM-a

Ako punu hijerarhijsku strukturu MQM-a zamislimo kao graf, možemo reći da svaki čvor (uključujući i one na najvišoj razini, poput točnosti i fluentnosti) može služiti kao vrsta problema u klasifikacijskoj shemi, dok djeca određenog čvora mogu predstavljati specifične slučajeve roditeljskog čvora. Zbog visoke podesivosti, MQM rješava problem jednodimenzionalnih, nepodesivih metrika poput LISA-inog modela. Nadalje, MQM standardizira terminologiju koja se koristi za klasificiranje problema, što znači da se metrike različitih korisnika mogu uspoređivati na mjestima gdje postoje preklapanja. Bitno je također naglasiti da je je MQM zamišljen kao neovisan o jeziku, zbog čega bi trebao biti primjenjiv na sve jezične parove.[33]

Višedimenzionalne metrike pretežno su namijenjene za analitičke metode evaluacije, tj. identificiranje specifičnih pogrešaka u prijevodu u svrhu njihove kvantifikacije. U takvim analitičkim metodama pogreške se smatraju odstupanjima u tekstu sukladno specifikacijama, što znači da se pogreške koje nisu dio klasifikacijske sheme ne prebrojavaju. Međutim, MQM se također može prilagoditi za holističku primjenu, tj. evaluaciju teksta kao cjeline. U tom slučaju vrste problema iz metrike pretvaraju se u pitanja o tekstu, a odgovori se pružaju pomoću popratnih

[33] Lommel, A.; Uszkoreit, H.; Burchardt, A. Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. Revista Tradumàtica: tecnologies de la traducció. 2014. Str. 457.

skalarnih vrijednosti. Na primjer, ako se metrika sastoji od točnosti, koja se dijeli na pogrešan prijevod i izostavljanje, i fluentnosti, koja se dijeli na gramatiku i stil, postavljaju se pitanja o tekstu na način prikazan u sljedećoj tablici.

	Ne	Djelomično	Da
Točnost			
[1a] Je li kakav sadržaj nepotrebno izostavljen iz prijevoda?			
[1b] Sadrži li tekst pogrešne prijevode koji mijenjaju značenje izvornog teksta?			
[1c] Postoje li kakve druge pogreške u točnosti koje utječu na prijevod?			
Fluentnost			
[2a] Postoje li kakvi gramatički problemi u prevedenom tekstu?			
[2b] Ima li stilističkih problema u prevedenom tekstu?			
[2c] Postoje li kakvi drugi problemi povezani s fluentnošću prevedenog teksta?			

Tablica 2: Holistička primjera MQM-a

U usporedbi s analitičkom metrikom, redoslijed problema kod holističke primjene MQM-a je obrnut: konkretnije vrste problema navode se ispred općenitijih (npr. stilistički problemi navedeni su ispred inače nadređene fluentnosti). Po svemu drugome holistička metrika preklapa se u potpunosti s analitičkom. Međutim, kod kompleksnijih metrika s mnogo vrsta problema, potrebna je određena doza pojednostavljivanja kada se analitička metrika pretvara u holističku. Na primjer, pitanja poput „Postoje li kakvi problemi kod uporabe funkcijskih riječi u prevedenom tekstu?“ najvjerojatnije nisu pretjerano informativna i korisna u holističkoj evaluaciji kada treba utvrditi zadovoljava li tekst specifikacije.

Primjer holističke primjene Višedimenzionalnih metrika možemo vidjeti i u TAUS-ovoj evaluaciji kvalitete, gdje se ocjenjuju fluentnost i točnost prijevoda uz pomoć skale vrijednosti koja je drugačija od one navedene u prethodnoj tablici. Budući da je jedan od dijelova testiranja kvalitete u Dinamičkom okviru kvantifikacija tipoloških pogrešaka koje su preuzete iz MQM-a, možemo zaključiti da su fluentnost i točnost isto tako preuzete kao dio Višedimenzionalnih metrika. To također znači da se rezultati dobiveni korištenjem ovih dviju različitih metoda donekle usporedivi.

5.1. Alati za anotaciju pomoću Višedimenzionalnih metrika

Nakon odabira metrike koja će se koristiti u određenom projektu, moguće je izabrati je izabrati jedan od dvaju različitih alata za anotaciju: MQM Scorecard ili translate5. Nažalost, u trenutku pisanja ovog rada, javni pristup alatu MQM Scorecard nije moguć zbog radova na njegovoj nadogradnji.[34] Nadalje, instalacijski paket za alat translate5 također još nije dovršen, zbog čega je moguće pristupiti samo njegovoj demonstracijskoj verziji. Bez obzira na to, i dalje je moguće pristupiti uputama i specifikacijama za korištenje pa će u nastavku rada biti opisani načini rada u oba alata.

5.1.1. MQM Scorecard

MQM Scorecard je jednostavan alat otvorenog koda koji omogućava anotaciju tekstova sravnjenih na razini segmenta pomoću MQM-ovih metrika. Ovaj alat zahtjeva prethodnu instalaciju na računalo, nakon koje mu je moguće pristupiti otvaranjem stranice editor.php koja se nalazi u instalacijskoj mapi. Za stvaranje novog projekta, potrebno je dodati „id=“ i neiskorišteni identifikacijski broj URL-u koji stranica otvara. Na primjer, ako se instalacija nalazi na *localhost* destonaciji u mapi /scorecard, URL izgleda ovako: http://localhost/scorecard/editor.php?id=123. Nakon toga otvara se sučelje u kojem još nije započet projekt i u kojem nisu dodane projektne datoteke. Projektne datoteke dodaju se klikom na karticu „Project Info“, gdje je potrebno postaviti bitext datoteku i XML datoteku s metrikama.[35]

U bitext datoteci nalaze se sravnjeni bitekstovi, tj. parovi segmenata na izvornom i ciljnom jeziku. Način segmentacije nije definiran zbog toga što Scorecard koristi bilo koju vrstu segmentacije koja se nalazi u datoteci. No datoteka za evaluaciju mora zadovoljavati sljedeće uvjete:

- Tekst mora biti kodiran u formatu UTF8 ili formatu ASCII s HTML kodovima za posebne znakove.
- Parovi segmenata moraju biti odijeljeni znakovima za novi red (*newline*), iako se unutar segmenta također može dodati oznaka
/
.

[34] Resources | QTLaunchpad. http://www.qt21.eu/launchpad/content/resources. 6.9.2016.
[35] Creating and annotating projects | Q Launchpad. http://www.qt21.eu/launchpad/node/1342#. 6.9.2016.

- Izvorni i ciljni segment odvajaju se tabulatorom. Segmenti ne smiju sadržavati znak tabulatora jer će biti nepravilno prikazani.
- Segmenti mogu sadržavati ograničene HTML oznake na istoj razini (*inline*). Podržane su sljedeće oznake: sa style atributom, ili <i>, ili <b>, <sub>, <sup> te <code>[36]

Primjer strukture bitext datoteke koja sadrži tri segmenta može se vidjeti u sljedećem okviru. Svaki par segmenata nalazi se u zasebnom redu, a izvorni i ciljni segment odvojeni su tabulatorom.

Action Sheet Nr. 1: MHPSS Core Principles	Akcija list br. 1: MHPSS Osnovna načela
Area	Područje
All event types, all target groups, all phases	Svi tipovi događaja, sve ciljne skupine, u svim fazama

XML datoteka s definicijom metrike jednaka je po strukturi onoj koja se koristi i u alatu translate5. U njoj je navedena shema svih problema koji se označavaju u anotatorskom zadatku. Imena problema mogu biti proizvoljna, ali projekt QTLaunchpad preporučuje da se koristi terminologija Višedimenzionalne metrike.[37] Datoteka sadrži korijenski element koji sadrži jedan ili više drugih elemenata. Elementi mogu biti prazni ili mogu sadržavati dodatne ugniježđene elemente. U sljedećem okviru nalazi se primjer strukture XML datoteke s 14 vrsta problema.

```
<issues>
	<issue type="Accuracy" level="0" display="yes">
		<issue type="Mistranslation" level="1" display="yes">
			<issue type="Terminology" level="2" display="yes" />
		</issue>
		<issue type="Omission" level="1" display="yes" />
		<issue type="Addition" level="1" display="yes" />
		<issue type="Untranslated" level="1" display="yes" />
	</issue>
	<issue type="Fluency" level="0" display="yes">
		<issue type="Content" level="1" display="no">
			<issue type="Register" level="2" display="yes" />
		</issue>
		<issue type="Mechanical" level="1" display="no">
			<issue type="Spelling" level="1" display="yes" />
			<issue type="Typography" level="1" display="yes" />
			<issue type="Grammar" level="1" display="yes" />
		</issue>
		<issue type="Unintelligible" level="1" display="yes" />
	</issue>
```

[36] Data formats | QTLaunchpad. http://www.qt21.eu/launchpad/node/1345#. 6.9.2016.
[37] Setting up a translate5 project | QTLaunchpad. http://www.qt21.eu/launchpad/node/1336. 6.9.2016.

Nakon učitavanja bitext i XML datoteke, može se početi s označavanjem. Slika sučelja alata nalazi se u prilogu zbog toga što bi zauzimala previše prostora u glavnom dijelu teksta. Na gornjoj strani sučelja nalaze se segmenti za evaluaciju, a na dnu se nalazi skupina tipaka koje se koriste za dodavanje problema segmentima. Segment koji treba anotirati aktivira se dvostrukim klikom na taj segment, nakon čega se oko njega pojavljuje crveni rub. Problemi u aktivnom segmentu dodaju se klikom na odgovarajuću vrstu problema u donjem dijelu sučelja te je pritom također potrebno za svaki problem navesti je li velik, mali ili kritičan. Bitno je napomenuti da se problemi iz kategorije „točnost“ mogu dodavati samo na ciljne segmente, dok se svi drugi problemi mogu dodavati na bilo koju stranu jezičnog para.

Za razliku od alata translate5, u kojem se problemi pridružuju određenom dijelu segmenta u kojem se nalaze, u alatu MQM Scorecard problemi se pridružuju segmentu kao cjelini. Međutim, u alatu Scorecard moguće je označavati dijelove teksta žutim markerom, čime se može ukazati na to gdje se problemi nalaze u segmentu. Osim toga, moguće je svakom segmentu dodavati bilješke ako za to postoji potreba.

5.1.2. translate5

Alat translate5 također je namijenjen anotaciji pomoću Višedimenzionalne metrike. Međutim, ovaj je alat malo kompleksniji od alata MQM Scorecard zato što podržava detaljniju anotaciju i napredne mogućnosti upravljanja korisnicima. Kao što je već spomenuto, najjednostavniji način pristupa alatu trenutno je putem demonstracijske verzije. Toj se verziji može pristupiti na web-adresi http://www.translate5.net/login pomoću predefiniranih korisničkih imena i lozinki koje se nalaze na stranici iznad okvira za prijavu. Problem s takvim pristupom je u tome što koristi dijeljenu verziju alata kojoj može pristupiti bilo tko i uređivati tuđe projekte. Stoga je za potrebe ovog rada korištena demo verzija alata koja se pokreće na lokalnom računalu.

Nakon prijave, moguće je započeti novi projekt klikom na gumb „Add Task“ u prozoru „Task Overview“. Pri stvaranju projekta, potrebno je navesti ime projekta, redni broj, izvorni i ciljni jezik te priložiti projektne datoteke. Također je moguće navesti prijelazni jezik, datum narudžbe i isporuke te broj riječi.

Kao i kod alata MQM Scorecard, ovdje je također potrebno priložiti dvije datoteke: datoteku u CSV formatu u kojoj se nalaze izvorni i ciljni segmenti te XML datoteku s metrikom,

koja je ista kao i za alat MQM Scorecard. U CSV datoteci trebaju se poštivati sljedeće specifikacije:

- Sadržaj svakog polja mora se nalaziti između dvostrukih navodnika.
- Svi drugi dvostruki navodnici u tekstu moraju biti prikazani kao par dvostrukih navodnika, ("").
- Polja moraju biti odvojena zarezima.
- Datoteka mora završavati oznakom kraja retka ili na kraju posljednjeg reda.
- Preporuča se korištenje UTF-8 kodne stranice.[38]

Za jednostavan projekt koji sadrži izvorni segment i ciljni segment koji je moguće uređivati, CSV datoteka treba sadržavati sljedeća polja (stupce):

- mid – tekstualni identifikator para segmenata, koji nije prikazan korisniku,
- source – izvorni tekst
- target – ciljni tekst

Za projekte s više stupaca koje je moguće uređivati, potrebna su samo polja *mid* i *source*, dok ostala polja mogu imati proizvoljna imena. U tom slučaju imena polja koja su prikazana u korisničkom sučelju navode se u prvom retku projektne datoteke. Primjer CSV datoteke u kojoj se nalaze 3 segmenta i izlazi dvaju sustava za strojno prevođenje može se vidjeti u idućem okviru.

```
“mid”,”source”, “GoogleTranslate”,“BingTranslator”

“2”,“Action Sheet Nr. 1: MHPSS Core Principles”,“Akcija list br. 1: MHPSS Osnovna načela”,“Akcija list Nr. 1: MHPSS osnovna načela”

“3”,“Area”,“područje”,“Područje”

“4”,"All event types, all target groups, all phases","Svi tipovi događaja, sve ciljne skupine, u svim fazama","Svih vrsta događaja, sve ciljane skupine, sve faze"
```

Projektne datoteke moraju se nalaziti u komprimiranoj datoteci u kojoj se nalazi direktorij „proofRead“. U tom direktoriju moguće je priložiti više od jedne CVS datoteke za anotiranje, dok se XML datoteka, koju je potrebno imenovati „QM_Subsegment_Issues.xml“, treba nalaziti u korijenskoj mapi. Ako XML datoteka nije priložena, sustav će koristiti metriku iz jezgre MQM-a.

[38] Setting up a translate5 project | QTLaunchpad. http://www.qt21.eu/launchpad/node/1336. 6.9.2016.

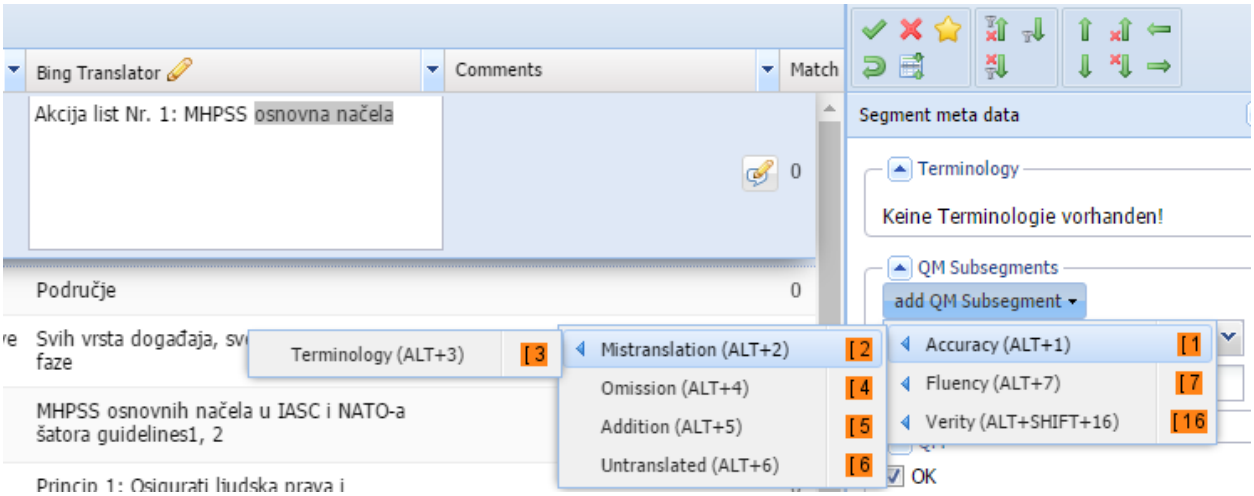
U slučaju odstupanja od strukture direktorija, pri postavljanju datoteke pojavit će se pogreška. Ako se priloži više CSV datoteka u direktoriju „proofRead“, prikazivat će se sve zajedno u korisničkom sučelju alata translate5, ali će biti prikazane kao poveznice koje omogućavaju korisniku prebacivanje na prvi segment svake datoteke.

Kada je projekt kreiran, nisu mu dodijeljeni korisnici, što znači da je njih potrebno naknadno dodavati. Za dodavanje postojećih korisnika potrebno je kliknuti na ikonu „Task-specific properties“ za željeni projekt. To otvara prozor gdje je klikom na gumb „Add user“ moguće dodati korisnike, kojima je potrebno navesti ime, ulogu i status. Za anotatorske zadatke kao ulogu korisnika potrebno je staviti „proofreader“ (lektor), a status treba biti „open“. U slučaju pogreške pri postavljanju korisnika, korisnike je uvijek moguće naknadno uređivati. Ako korisnik još ne postoji, moguće je dodati nove korisnike klikom na gumb „Add user“. To otvara malo kompleksniji dijaloški okvir u kojem treba navesti ime i prezime korisnika, spol, adresu elektroničke pošte, moguće uloge (urednik i voditelj projekta) te korisničko ime i lozinku. Korisnici koji su postavljeni kao urednici mogu anotirati i uređivati projekte, dok voditelji projekta (Pm) mogu kreirati i modificirati projekte te dodavati nove korisnike i upravljati njima (uređivati njihove podatke, brisati ih, mijenjati njihove lozinke itd.).[39]

Kad je projekt postavljeni i anotatori su zadani, korisnici mogu pristupiti alatu i početi označavati. Nakon prijave, korisnici mogu vidjeti popis svih projekata koji su im dostupni i otvoriti željeni projekt. Korisničko sučelje za vrijeme anotacije sastoji se od popisa segmenata i uređivača s lijeve strane te okvira za dodavanje metapodataka s desne. Segmenti mogu biti prikazani u tri različita pogleda: pregledavačkom (*View*), uređivačkom (*Editing*) i ergonomičnom (*Ergonomic*). U pregledavačkom pogledu prikazani su stupci koji prikazuju broj segmenta, status, izvorni tekst, jedan ili više prijevoda, komentare, stopu podudaranja i zadnjeg anotatora, dok su u uređivačkom pogledu polja sa ciljnim segmentima prisutna dvaput za svaki prijevod; u jednom ih je polju moguće uređivati, a u drugom je polju prikazan ciljni segment u svom izvornom obliku. U ergonomičnom pogledu prikazani su samo izvorni segment, jedan od ciljnih segmenata te broj i status segmenta. Međutim, ovaj pogled također omogućava uređivanje segmenata. Detaljnija slika sučelja prikazana je na slici koja se nalazi u poglavlju Prilozi.

[39] User management | QTLaunchpad. http://www.qt21.eu/launchpad/node/1338. 6.9.2016.

Segmenti koje je moguće uređivati otvaraju se dvostrukim klikom te se u isto vrijeme otvara i okvir za dodavanje metapodataka. Kako bi se dodala oznaka MQM-a na određeni dio teksta, potrebno je označiti taj dio teksta u segmentu i dodati vrstu problema s desne strane sučelja, klikom na gumb „add QM Subsegment“ i odabirom odgovarajuće vrste problema. Alat translate5 dopušta preklapanje raspona oznake pa nije potrebno voditi brigu o granici svake oznake.[40]



Slika 6: Dodavanje problema dijelu segmenta u alatu translate5

Odabirom vrste problema taj se problem dodaje označenom rasponu u segmentu. Oznaka problema prikazana je u uglatim zagradama u obliku numeričke oznake koja odgovara broju vrste problema na popisu problema u izborniku. Pomoću skupine kontrolnih gumba koja se nalazi u okviru s desne strane alata moguće je prihvaćati ili odbijati oznake te se kretati kroz datoteku. Međutim, korištenje tih gumba nije neophodno jer se oznake automatski spremaju klikom na sljedeći segment.

Ako je za potrebe anotatorskog zadatka potrebno označavati težinu problema, to se može učiniti klikom na padajući izbornik koji nudi tri različite težine: kritični, mali i veliki problem. Trenutna verzija alata translate5 ne omogućava postavljanje težine problema nakon što je oznaka već dodana pa se težina problema mora dodati prije odabira vrste problema. To predstavlja problem ako je težinu problema potrebno promijeniti ili naknadno dodati zbog toga što u tom slučaju prvo treba obrisati oznaku, namjestiti težinu i ponovno dodati oznaku.

Alat translate5 također omogućava dodavanje komentara na segmente i na same oznake. Komentari na oznake dodaju se unošenjem teksta u polje „Comment“ prije odabira vrste problema

[40] Logging in and annotating | QTLaunchpad. http://www.qt21.eu/launchpad/node/1339. 6.9.2016.

jer ih nije moguće naknadno dodavati. Komentar na segment kao cjelinu dodaje se odabirom segmenta i klikom na ikonu „Add comment“. Svi se komentari spremaju uz korisničko ime komentatora i vremensku oznaku koja pokazuje kada je komentar napisan.

Nakon završetka označavanja, zadatak je moguće završiti na dva načina. Prvi način uključuje povratak na popis zadataka klikom na gumb Task i odabirom opcije „Return to task list“. Drugi način za završetak zadatka je klikom na gumb „End Task“, nakon čega je zadatak označen kao završen i šalje se notifikacija administratoru projekta. Međutim, odabirom ove opcije zadatak se zaključava i korisnici ga više ne mogu uređivati osim ako administrator ponovno otvori projekt. Zbog toga je preporučeno da se zadaci ne završavaju na taj način, nego da se korisnici samo vrate na popis zadataka i osobno obavijeste voditelja projekta da je zadatak završen.[41]

5.1.3. Smjernice za označavanje pomoću MQM-a

Budući da je označavanje pomoću Višedimenzionalne metrike kompleksan zadatak, anotatori mogu naići na određene nejasnoće pri označavanju. Zbog toga je za svaki anotatorski zadatak potrebno pripremiti smjernice kojima će se anotatori koristiti kao referentnim materijalima. Osim smjernica, preporuča se i izrada stabla odlučivanja kojim se anotatori mogu služiti kako bi naučili pravilno pridruživati oznake, ali i kako bi riješili potencijalne nedoumice koje se mogu pojaviti. Primjer stabla odlučivanja kakvo je primjenjivo na korištenje metrike iz jezgre MQM-a nalazi se u prilozima. Tom se stablu pristupa tako da se počne od gornjeg lijevog okvira i odgovara se na pitanja kako bi se odabrala ispravna vrsta problema. Hijerarhija problema u stablu razlikuje se od hijerarhije problema u alatu translate5 zato što se u stablu najprije eliminiraju specifične vrste problema, a tek se tada potencijalno bira generalizirana vrsta problema.

Kao primjer smjernica koje su korisne za većinu anotatorskih zadataka u alatu translate5 možemo uzeti *Praktične smjernice za korištenje MQM-a u znanstvenim istraživanjima o kvaliteti prijevoda*.[42] U tim se smjernicama najprije definira što je to pogreška: bilo koji problem u prijevodu koji ne odgovara izvornom tekstu ili se smatra nepravilnim u ciljnom jeziku. Lista problema dijeli se na dvije glavne kategorije, a to su točnost i fluentnost, od kojih svaka sadrži detaljnije potkategorije. Preporučeno je da se u anotatorskim zadacima uvijek koristi ispravna potkategorija umjesto generičke kategorije problema. Međutim, ako nije jasno koja je kategorija

[41] Logging in and annotating | QTLaunchpad. http://www.qt21.eu/launchpad/node/1339. 6.9.2016.
[42] Burchardt, A.; Lommel A. Practical Guidelines for the Use of MQM in Scientific Research on Translation Quality. 2014.

najprikladnija, bolje je ne pogađati. U tom se slučaju preporuča odabrati kategoriju na razini za koju je moguće najpreciznije odrediti da odgovara problemu u tekstu.

U spomenutim smjernicama navedeno je nekoliko glavnih pravila kojih se treba pridržavati pri označavanju. Ta su pravila primjenjiva na sve vrste anotacije, bez obzira na metriku.

1. Ako se više vrsta problema iz metrike može pridružiti problemu u tekstu (npr. slaganje riječi, oblik riječi, gramatika, fluentnost), potrebno je odabrati prvu vrstu problema na koju upućuje stablo odlučivanja. To je stablo organizirano prema sljedećim načelima:
    a. stablo preferira specifičnije vrste problema (npr. vrsta riječi) pred općenitijima (npr. gramatika). Međutim, ako specifična vrsta problema nije primjenjiva, anotatori su upućeni na korištenje općenite vrste problema.
    b. Općenite vrste problema koriste se kada se pojave problemi općenite prirode ili kada za specifični problem ne postoji precizno definirana vrsta problema u metrici.
2. Manje je više. Potrebno je označiti samo relevantan dio teksta. Na primjer, ako je u frazi pogrešno napisana ili prevedena samo jedna riječ, označava se samo ta riječ, a ne cijela fraza. Ako se greška pojavljuje u dvije riječi između kojih se pojavljuju neke druge riječi, treba odvojeno označiti samo te dvije riječi.
3. Ako uklanjanje jedne pogreške može popraviti sve druge, treba označiti samo tu pogrešku. Na primjer, ako uklanjanje pogreške u slaganju riječi popravlja probleme koji nastaju zbog te pogreške, potrebno je označiti samo tu pogrešku u slaganju riječi, ali ne i pogreške koje su nastale kao rezultat te pogreške.
4. Ako se u jednoj riječi pojavljuje više pogrešaka (npr. jedna riječ sadrži pogrešku u pravopisu, ali osim toga predstavlja i nepotrebnu funkcijsku riječ), potrebno je odvojeno unijeti obje vrste greške i označiti istu riječ u oba slučaja.
5. Ako nije očito koju vrstu problema treba odabrati, preporučeno je odabrati općenitiju kategoriju. Kategorije točnost i fluentnost mogu se koristiti samo kada je nemoguće odrediti narav problema.

Osim ovih glavnih pravila, u *Smjernicama* se nalaze i neki primjeri problematičnih slučajeva koji su se pojavili u praksi.

- **Funkcijske riječi**: U nekim slučajevima problemi povezani s funkcijskim riječima ne odgovaraju podjeli iz stabla odlučivanja zato što se nalaze u kategoriji fluentnosti, iako utječu na značenje. Usprkos tome, preporuča se da se uvijek koristi prikladna vrsta problema iz kategorije funkcijskih riječi.
- **Redoslijed riječi**: Problemi s redoslijedom riječi često obuhvaćaju dugačke raspone u tekstu. Kada se takvi problemi označavaju, potrebno je označiti najmanji mogući dio teksta koji je potrebno pomaknuti kako bi se popravio problem.
- **Riječi pisane s crticom**: Problemi s riječima koje se pišu s crticom mogu se pojaviti u neprevedenim riječima te se u tom slučaju trebaju označiti kao vrsta pogreške koja se odnosi na neprevedene riječi. U drugim slučajevima, taj tip pogreške spada pod pravopis.
- **Broj** (jednina i množina): Slaganje riječi u broju označava se kao pogrešan prijevod.
- **Terminologija**: Ako se u prijevodu rabi pogrešan termin, tu pogrešku treba označiti kao pogrešku u terminologiji, bez obzira na to je li tekst i dalje razumljiv. U anotatorskim zadacima gdje je terminologija uključena kao vrsta problema potrebno je također priložiti glosar termina ako termini nisu općepoznati anotatoru.
- **Nerazumljivo**: kategorija „nerazumljivo“ koristi se ako je tekst nemoguće razumjeti, a razlog tomu ne može se analizirati pomoću stabla odlučivanja. Ta se kategorija koristi samo kao posljednja opcija u tekstovima gdje narav problema nije nimalo jasna.
- **Slaganje**: Kategorija slaganja općenito se odnosi na slaganje između subjekta i predikata i slaganje u rodu, broju i padežu.
- **Neprevedeno**: Mnoge riječi mogu izgledati kao da su prevedene, a da nakon toga nisu primijenjena ispravna pravila o pisanju velikog i malog slova ili pisanju riječi s crticom. Međutim, ako je riječ ili fraza identična riječi ili frazi iz izvornog teksta, treba ju tretirati kao neprevedenu, bez obzira na to može li se problem također pripisati grešci u pravopisu.
- Općenito se preporuča da se pogreške označavaju s najmanjim mogućim rasponom. Oznaka mora obuhvaćati samo dio teksta koji je potreban da bi se odredio problem. U nekim slučajevima to znači da je potrebno označiti dva različita raspona kako bi se odredila jedna pogreška.

Uz pravila, u *Smjernicama* je također definirano što koja vrsta greške iz Višedimenzionalne metrike označava. Te su definicije i hijerarhijska struktura problema prikazani su u nastavku.

- **Točnost.** Točnost pokazuje do koje mjere ciljni tekst ispravno prenosi značenje sadržano u izvornom tekstu.
    - **Pogrešan prijevod.** Sadržaj ciljnog teksta pogrešno predstavlja sadržaj izvornog teksta.
        - **Terminologija.** Pogrešno su prevedeni termini svojstveni nekom području ili industriji.
    - **Izostavljanje.** Sadržaj koji postoji u izvornom tekstu ne pojavljuje se u ciljnom tekstu.
    - **Dodavanje.** Ciljni tekst sadrži dijelove koji ne postoje u izvornom.
    - **Neprevedeno.** Sadržaj koji je moguće prevesti nije preveden.
- **Fluentnost.** Fluentnost se odnosi na jednojezične značajke izvornog ili ciljnog teksta, ovisno o dogovorenim specifikacijama, ali neovisno o odnosu između izvornog i ciljnog teksta. Drugim riječima, problemi s fluentnošću mogu se uočiti bez obzira na to radi li se o prijevodu ili o nekom drugom tekstu.
    - **Pravopis.** Pojavljuju se problemi povezni s pravilnim pisanjem riječi (uključujući velika i mala slova)
    - **Tipografija.** Postoje problemi povezani s mehaničkim prikazom teksta. Ova se kategorija koristi za sve tipografske pogreške osim za one obuhvaćene kategorijom pravopisa (npr. točke, zarezi itd.)
    - **Gramatika.** Pojavljuju se problemi povezani s gramatikom i sintaksom u tekstu, koji nemaju veze s pravopisom.
        - **Oblik riječi.** Korišten je pogrešan oblik riječi. Potrebno je birati podtipove ove kategorije kad god je to moguće.
            - **Vrsta riječi.** Iskorištena je pogrešna vrsta riječi s korijenom ispravne riječi.
            - **Slaganje.** Dvije ili više riječi ne slažu se u padežu, broju, licu ili nekoj drugoj gramatičkoj značajci.
            - **Glagolski oblik.** Iskorišten je pogrešan glagolski oblik s obzirom na kontekst.

- **Redoslijed riječi.** Redoslijed riječi je pogrešan.
- **Funkcijske riječi.** Neispravno su korištene funkcijske riječi, poput prijedloga čestica i zamjenica.
    - **Nepotrebno.** Funkcijska riječ koja se pojavljuje u tekstu nije potrebna.
    - **Nedostaje.** U tekstu nedostaje funkcijska riječ koja bi se trebala pojaviti.
    - **Netočno.** Iskorištena je nepravilna funkcijska riječ.
- **Nerazumljivo.** Nemoguće je odrediti točnu narav pogreške koja predstavlja veliki problem u fluentnosti.

5.1.4. Mehanizam ocjenjivanja

Nakon označavanja pomoću Višedimenzionalne metrike, moguće je izračunati ocjenu označenog teksta. Ta se ocjena računa prema sljedećoj formuli neovisno o alatu.[43]

$$TQ = 100 - AP - (FP_T - FP_S) - (VP_T - VP_S)$$

Pri čemu je:

- TQ = ocjena kvalitete
- AP = težinski zbroj negativnih bodova pridruženih grani točnosti (*accuracy*)
- FP_T = težinski zbroj negativnih bodova koji su pridruženi grani fluentnosti (*fluency*) u ciljnom tekstu.
- FP_S = težinski zbroj negativnih bodova koji su pridruženi fluentnosti u izvornom tekstu. Ako se ciljni tekst ne ocjenjuje, $FP_S = 0$
- VT_P = težinski zbroj negativnih bodova koji su pridruženi grani istinitosti (*verity*) u ciljnom tekstu.
- VP_S = težinski zbroj negativnih bodova koji su pridruženi grani istinitosti u izvornom tekstu. Ako se izvornik ne ocjenjuje, $VP_S = 0$.

[43] Lommel, A. Multidimensional Quality Metrics. Meta Forum 2013. Prezentacija. 2013.

Negativni bodovi računaju se prema sljedećoj formuli.

$$P = \frac{(problemi_{mali} + 5 \times problemi_{veliki} + 10 \times problemi_{kritični})}{broj\ riječi}$$

Pri čemu je:

- P = negativni bodovi
- $problemi_{mali}$ = zbroj problema koji ne utječu na razumijevanje
- $problemi_{veliki}$ = zbroj problema koji utječu na razumijevanje teksta, ali ga ne čine beskorisnim
- $problemi_{kritični}$ = zbroj problema koji čine tekst beskorisnim
- broj riječi = ukupan broj riječi u uzorku

Negativni bodovi povezani s težinom problema mogu se prilagoditi posebnim potrebama, ali se zbog bolje operabilnosti i kompatibilnosti među sustavima preporuča korištenje pretpostavljenog modela za ocjenjivanje.[44] Nadalje, alati za anotaciju pomoću Višedimenzionalne metrike ne računaju ocjenu automatski, već ju je potrebno samostalno izračunati. Međutim, za taj zadatak također je moguće koristiti se tablicom za izračunavanje ocjene dostupnom na sljedećoj adresi: http://jira.translate5.net/secure/attachment/10002/scoreCardCore.xlsx. (7.9.2016.)

[44] Creating a translation quality score with MQM. http://www.qt21.eu/launchpad/node/1332. 6.9.2016.

6. Istraživanje

U svrhu testiranja opisanih alata provedeno je istraživanje na englesko-hrvatskom jezičnom paru u kojemu se koristila većina alata i njihovih funkcija. Primarni cilj istraživanja bio je testiranje jesu li rezultati koje pružaju DQF alati i alat translate5 usporedivi i komplementarni, s tezom da DQF alati mogu pružiti općeniti uvid u razinu kvalitete prijevoda, dok alat translate5 omogućava temeljitiju analizu i odgovor na pitanje *zašto* je neki prijevod bolji od drugoga. Drugi cilj istraživanja bio je usporedba dvaju sustava za strojno prevođenje i njihove učinkovitosti na različitim vrstama tekstova.

6.1 Povezana istraživanja

Napravljena su brojna slična istraživanja u kojima se vrši ljudska i/ili automatska evaluacija strojnih prijevoda s englesko-hrvatskim jezičnim parom. Seljan, Vičić i Brkić (2012)[45] provode istraživanje u kojem evaluiraju prijevode tekstova iz područja zakonodavstva s engleskog na hrvatski, prevedene pomoću besplatnog alata Google Translate dostupnog na internetu. Testni uzorak uključuje 200 rečenica, koje se dijele na duge i kratke rečenice. U tom se istraživanju provodi ljudska evaluacija, koja uključuje kriterije adekvatnosti i fluentnosti te analizu pogrešaka (neprevedene riječi, izostavljene riječi, nepotrebno prevedene riječi, morfološke pogreške, leksičke pogreške, sintaktičke pogreške i nepravilna interpunkcija). Osim toga, računa se i evaluacijska metrika BLEU i različite vrste korelacije. Testiranje adekvatnosti i fluentnosti mjeri se na skali od 1 do 5, te je ostvaren prosječni rezultat od 3,48 za kratke rečenice i 3,00 za duge rečenice. Rezultat metrike BLEU iznosi 0,25 za kratke rečenice i 0,20 za duge uz jedan set referentnih prijevoda, tj. 0,32 i 0,26 uz tri seta referentnih prijevoda. Osim toga, korelacijom između ljudske evaluacije i različitih vrsta pogrešaka pokazalo se da na fluentnost najčešće utječu morfološke pogreške, nakon čega slijede neprevedene i izostavljene riječi, dok na adekvatnost najviše utječu leksičke greške te neprevedene i izostavljene riječi.

[45] Seljan, S.; Vičić, T.; Brkić, M. BLEU Evaluation of Machine-Translated English-Croatian Legislation. // Proceedings of the Eighth International Conference on Language Resources and Evaluation. Istanbul, Turkey: European Language Resources Association, 2012.

Seljan i Dunđer (2015)[46] računaju automatske metrike (BLEU, NIST, METEOR i GTM) za strojne prijevode s engleskog i ruskog jezika na hrvatski. Njihov se uzorak sastoji od sveukupno 400 rečenica iz područja opisa grada te sadrži 100 rečenica za svaki jezični par, a prijevodi koji se evaluiraju predstavljaju izlaze *online* sustava za strojno prevođenje Google Translate i Yandex.Translate. U rezultatima ovog istraživanja pokazuje se da je rusko-hrvatski jezični par dobio bolje ocjene od englesko-hrvatskog za oba alata. Pri uspoređivanju rezultata dvaju alata, Yandex.Translate postigao je malo bolje rezultate, u rasponu od 0,2 do 13 %. Google Translate ostvario je bolje rezultate na hrvatsko-engleskom jezičnom paru, dok je Yandex.Translate ostvario bolje rezultate za rusko-hrvatski par u svim automatskim metrikama.

Brkić, Vičić i Seljan (2009)[47] evaluiraju prijevode alata Google Translate s engleskog na hrvatski. U tom se istraživanju evaluiraju prijevodi tri vrste teksta: tekst o korpusnoj lingvistici, tekst o Vladinom planu reforme i tekst o perilici za posuđe. Šest evaluatora vrši manualnu evaluaciju uzorka od 21 rečenice te ocjenjuju fluentnost i adekvatnost na skali od 1 do 5. Pritom postižu prosječni rezultat 2,98 za fluentnost i 3,36 za adekvatnost. Standardna devijacija na razini pitanja iznosi između 0,2 i 1,03 za fluentnost i između 0,41 i 1,05 za adekvatnost, dok standardna devijacija na razini evaluatora iznosi između 0,6 i 1,06 za fluentnost te između 0,6 i 1,32 za adekvatnost. Osim toga, proveden je Hi-kvadrat test, kojim se pokazalo da ne postoji značajna razlika u pridruživanju ocjene 3 među evaluatorima, kako za kriterij fluentnosti, tako i za kriterij adekvatnosti. Analizom rezultata tog testa također se pokazalo da polovica evaluatora smatra kriterije fluentnosti i adekvatnosti usko povezanima što se tiče ocjene 3, dok je druga polovica sposobna bolje razlikovati te kriterije i ocjenjivati ih drugačije.

[46] Seljan, Sanja; Dunđer, Ivan. Machine Translation and Automatic Evaluation of English/Russian-Croatian // Proceedings of the International Conference "Corpus Linguistics - 2015". St. Petersburg, Rusija : St. Petersburg State University, 2015. 72-79

[47] Brkić, Marija; Vičić, Tomislav; Seljan, Sanja. Evaluation of the Statistical Machine Translation Service for Croatian-English. // 2nd international conference: The future of information sciences (INFuture 2009) : Digital resources and knowledge sharing : proceedings / Stančić, Hrvoje ; Seljan, Sanja ; Bawden, David ; Lasić-Lazić, Jadranka ; Slavić, Aida (ur.). Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009. 319-332

Seljan, Brkić i Kučiš (2011)[48] evaluiraju tekstove iz četiriju domena (opis grada, pravo, nogomet i monitori) prevedene s hrvatskog na engleski jezik pomoću alata Google Translate, Stars21, InterTran i Translation Guide te tekstove prevedene s engleskog na hrvatski pomoću alata Google Translate. Vrsta evaluacije provedena u ovom istraživanju ocjenjivanje je fluentnosti i adekvatnosti na skali od 1 do 5. Međutim, glavni je cilj ovog istraživanja računanje koeficijenta Fleiss kappa kako bi se dobio uvid u slaganje među evaluatorima. Taj koeficijent pokazuje značajno, a ponekad i savršeno slaganje u evaluaciji četiriju prijevodnih servisa. Gotovo savršeno slaganje pokazalo se u ocjenjivanju alata Translation Guide kao najgoreg prijevodnog servisa. Značajno slaganje pokazalo se u evaluaciji alata Stars21 i Google Translate, koji su dobili najviše ocjene. Za evaluaciju alata InterTran, koji je ocjenjen malo bolje od alata Translation Guide, pokazalo se umjereno slaganje među evaluatorima. Osim evaluacije alata i računanja slaganja među evaluatorima, napravljena je i analiza grešaka, kojom se pokazalo da su na ocjene prijevoda najviše utjecale neprevedene riječi, dok su se morfološke i sintaktičke greške u manjoj mjeri odražavale na ocjene.

Brkić, Seljan i Vičić (2013)[49] provode automatsku i ljudsku evaluaciju prijevoda teksta iz područja prava s engleskog na hrvatski. Ljudska evaluacija u tom istraživanju uključuje ocjenjivanje fluentnosti i adekvatnosti te analizu grešaka, dok se u automatskoj evaluaciji rabe metrike BLEU, NIST, F-mjera i WER. Testni set uključuje 100 rečenica iz područja prava te 100 rečenica iz područja religije, psihologije, obrazovanja itd. prevedenih pomoću alata Google Translate. Rezultati ljudske evaluacije za prvi testni set iznose 3,03 za fluentnost i 3,04 za adekvatnost, dok za drugi testni set iznose 3,3 za fluentnost i 3,67 za adekvatnost. U analizi pogrešaka, proučavaju se sljedeće kategorije: neprevedene/izostavljene riječi, nepotrebne riječi u prijevodu, sintaktičke greške – redoslijed riječi i pravopisne greške. U rezultatima analize vidi se da je u prijevodima prisutan najveći broj morfoloških pogrešaka, dok su druge vrste pogrešaka manje zastupljene. Rezultati automatskih metrika koriste se za proučavanje kako predobrada, tj opojavnjičivanje, pretvaranje svih slova u mala slova i uklanjanje interpunkcije, utječe na rezultate.

[48] Seljan, Sanja; Brkić, Marija; Kučiš, Vlasta. Evaluation of Free Online Machine Translations for Croatian-English and English-Croatian Language Pairs. // INFuture2011: The Future of Information Sciences - Information Sciences and e-Society / Billenness, Clive ; Hemera, Annette ; Mateljan, Vladimir ; Banek Zorica, Mihaela ; Stančić, Hrvoje ; Seljan, Sanja (ur.). Zagreb : Department of Information Sciences, 2011.. Str. 331-344

[49] Brkić, Marija; Seljan, Sanja; Vičić, Tomislav. Automatic and Human Evaluation on English-Croatian Legislative Test Set. // Lecture Notes in Computer Science - LNCS. 7816 (2013) , 1; 311-317

Brkić, Seljan i Matetić (2011)[50] provode istraživanje u kojem evaluiraju strojne prijevode s engleskog jezika na hrvatski i s hrvatskog na engleski. Za prijevode na hrvatski koriste se četirima *online* sustavima za strojno prevođenje: alatima Google Translate, Stars21, Translation Guide i InterTran, dok se za prijevod na engleski jezik koriste samo alatom Google Translate. U evaluaciji se koriste trima automatskim metrikama: F-mjerom, BLEU-om i NIST-om, a osim toga provode i ljudsku evaluaciju, u koju je uključeno ocjenjivanje fluentnosti i adekvatnosti. Osim toga, u istraživanju se također proučava i korelacija između ljudskih i automatskih metrika, koja se na ovom uzorku nije pokazala znatnom.

6.2 Eksperimentalno istraživanje

Provedeno je istraživanje na strojnim prijevodima dvaju tekstova s engleskog na hrvatski jezik. Svaki tekst sadrži 50 segmenata, odnosno oko 700 riječi. Jedan od tih tekstova dolazi iz područja psihosocijalne pomoći u slučaju katastrofe, dok drugi tekst govori o ustrojstvu Europske unije. Kako bi rezultati među alatima bili što više usporedivi, bilo je bitno da se u svakom alatu evaluira ista količina teksta. Za prijevod tekstova korišteni su alati Google Translate i Bing Translator. Strojni prijevodi evaluirani su u DQF-ovom alatu za brzo uspoređivanje i alatu za evaluaciju kvalitete, gdje se ocjenjivala fluentnost, adekvatnost i tipološke pogreške, te su nakon toga označeni u alatu translate5 pomoću Višedimenzionalne metrike.

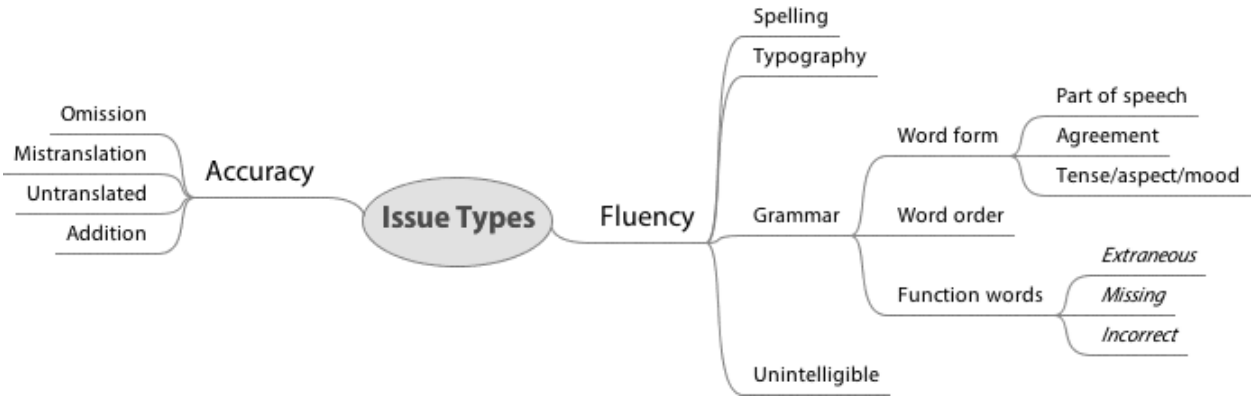
Zadatak evaluacije brzom usporedbom za obje vrste teksta obavilo je četiri evaluatora te je ostvaren koeficijent Fleiss Kappa od 0,6923 za područje Europske unije i 0,6588 za područje psihosocijalne pomoći. Prema tablici Landisa i Kocha[51] za interpretaciju tog koeficijenta, slaganje među evaluatorima možemo ocijeniti kao znatno. Svi su evaluatori bili kompetentni za izvršavanje ovog zadatka zbog njihove pozadine u lingvistici i odličnog poznavanja izvornog i ciljnog jezika. Na druga dva zadatka radio je samo jedan evaluator jer je njihova kompleksnost zahtijevala višu razinu stručnosti i iskustva kakvu nije bilo moguće ostvariti kod drugih evaluatora.

Od tipoloških pogrešaka iz DQF alata za evaluaciju kvalitete prebrojavale su se samo pogreške u fluentnosti i adekvatnosti, što znači da je izostavljena evaluacija stila, terminologije i

[50] Brkić, Marija; Seljan, Sanja; Matetić, Maja. Machine Translation Evaluation for Croatian-English and English-Croatian Language Pairs. // Proceedings fo the 8th International NLPCS Workshop: Human-Machine Interaction in Translation / Sharp, Bernardette ; Zock, Michael ; Carl, Michael ; Jakobsen, Arnt Lykke (ur.). Copenhagen : Copenhagen Business School, 2011. 93-104

[51] The measurement of observer agreement for categorical data. // Biometrics / Landis, J. R.; Koch, G. G. Washington DC : The International Biometric Society. 1977. Str. 159-174.

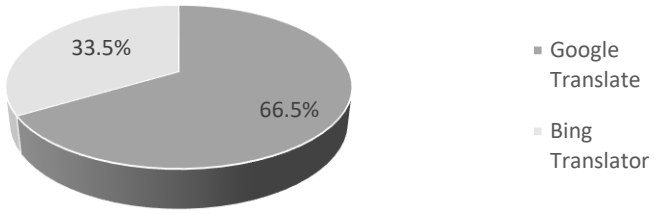
primjerenosti konvencijama regionalne sheme. To je učinjeno zato što se te kategorije obično ne koriste za dijagnostiku strojnih prijevoda u sklopu Višedimenzionalne metrike prilagođene za tu svrhu. Metrika koja je korištena za anotaciju u alatu *translate5* prikazana je na sljedećoj slici.



Slika 7: Višedimenzionalna metrika za dijagnostiku strojnih prijevoda

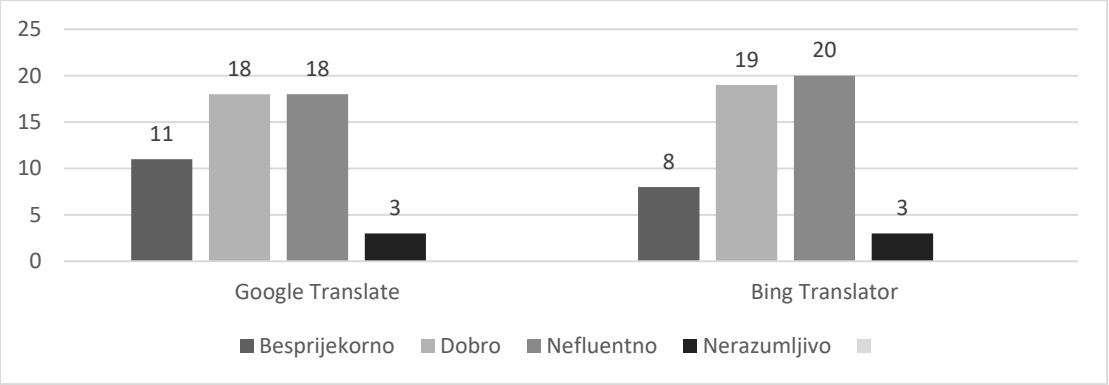
6.1 Rezultati istraživanja – psihosocijalna pomoć

Rezultati istraživanja bit će prikazani najprije za svaku vrstu teksta zasebno, nakon čega će se rezultati usporediti s obzirom na vrste tekstova. Za područje psihosocijalne pomoći, svih četiri evaluatora odabralo je da je u preko 60% slučajeva izlaz alata Google Translate bolji od izlaza alata Bing Translator. Rezultati svih evaluatora prilično su slični, a sveukupan omjer iznosi 68 naprama 32 posto, kao što je prikazano u sljedećem grafikonu.

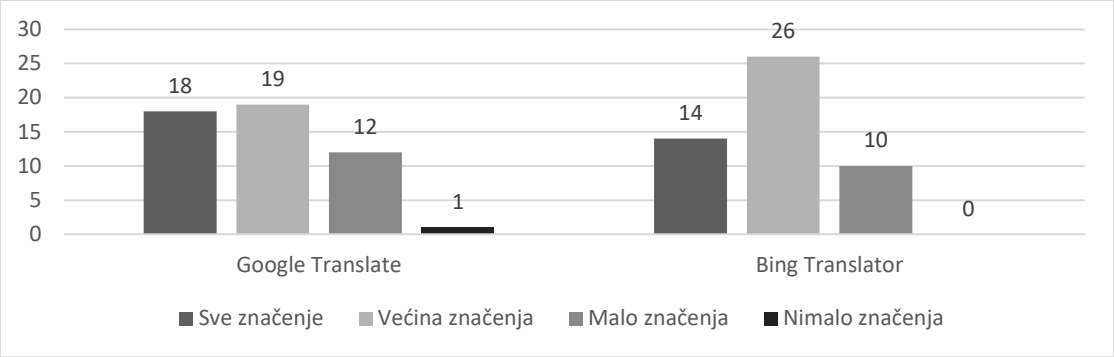


Grafikon 1: Preferirani izlaz alata - Psihosocijalna pomoć

U evaluaciji kvalitete, fluentnost izlaza alata Google Translate ocjenjena je prosječnom ocjenom 2.5 od 4, dok je za alat *Bing Translator* ocjena fluentnosti 2. To znači da je jezik izlaza oba alata okarakteriziran kao nefluentan, iako je izlaz alata Google Translate malo tečniji, s većim brojem besprijekornih segmenata i manjim brojem nefluentnih segmenata. Ocjena adekvatnosti jednaka je za oba alata te iznosi 3 od 4, što znači da je značenje izvornog teksta većinom preneseno u ciljni tekst. Međutim, proučavajući distribuciju adekvatnosti za oba alata može se primijetiti da je, iako je ocjena ista, Google Translate proizveo veći broj segmenata u kojima je zadržano sve značenje, dok Bing Translator ima značajno veći broj segmenata u kojima je zadržana samo većina značenja. Distribucija fluentnosti i adekvatnosti za oba alata prikazana je na sljedećim grafikonima.



Grafikon 2: Distribucija fluentnosti



Grafikon 3: Distribucija adekvatnosti

Rezultati prebrojavanja pogrešaka u DQF-ovu alatu za evaluaciju kvalitete nisu pružili puno bolji uvid u konkretne probleme prisutne u prijevodima teksta o psihosocijalnoj pomoći. Sveukupni broj pogrešaka za svaki alat razlikuje se samo u 3 pogreške više u prijevodu alata Bing Translator. Iako je alat Google Translate proizveo malo više pogrešaka u točnosti, a Bing Translator malo više pogrešaka u fluentnosti, razlike nisu dovoljno značajne da bi se mogli donositi zaključci. Međutim, budući da drugi testovi indiciraju da su izlazi alata Google Translate većinom bolji, možemo pretpostaviti da se u ovom istraživanju ova vrsta evaluacije nije isplatila. Kao najveći problem ove metode možemo prepoznati činjenicu da se pogreške samo prebrojavaju, dok ne postoji osjetljivost na težinu pogreške. Konkretni brojevi pogrešaka u prijevodu teksta o psihosocijalnoj pomoći prikazani su sljedećoj tablici.

	Točnost	**Fluentnost**	**Sveukupno**
Google	48	97	145
Bing	44	104	148

Tablica 3: Broj pogrešaka u tekstu o psihosocijalnoj pomoći

Označavanjem pogrešaka pomoću Višedimenzionalne metrike u alatu *translate5* napravljena je dublja analiza. S obzirom na to da u tom alatu postoji mogućnost pridruživanja težine svakoj pogrešci, moguće je dobiti realniju sliku o tome koji alat za strojno prevođenje proizvodi tekstove bolje fluentnosti i točnosti. Budući da se označavanje konkretnih riječi u prijevodu razlikuje od običnog prebrojavanja pogrešaka, broj pogrešaka označenih pomoću Višedimenzionalnih metrika razlikuje se od broja pogrešaka u DQF-ovoj evaluaciji kvalitete, iako su krovne kategorije pogrešaka iste. Osim toga, Višedimenzionalna metrika pruža mnogo zrnatije tipove pogrešaka, što nam omogućuje uvid u konkretne pogreške u prijevodima.

U prijevodu alata Google Translate označeno je sveukupno 178 pogrešaka, od čega se 56 pogrešaka odnosi na točnost, a 122 pogreške na fluentnost. Najveći broj pogrešaka imaju veliku težinu (137), dok malih pogrešaka ima 18 i kritičnih 23. Većina pogrešaka u točnosti (42) odnosi se na pogrešan prijevod, a samo su dvije pogreške pridružene neprevedenim riječima. Međutim, najveći broj kritičnih pogrešaka (14) pripada upravo kategoriji pogrešnog prijevoda, što uvelike smanjuje ocjenu točnosti. Što se tiče fluentnosti, najčešće se pogreške pojavljuju u slaganju riječi u rodu, broju i padežu (70), nakon čega slijede pogreške u pogrešnom odabiru vrste riječi (13). Najmanji se broj pogrešaka pojavljuje u tipografiji i funkcijskim riječima, s tim da je pogreškama u tipografiji i pravopisu najčešće pridodana mala težina.

Ocjena točnosti prijevoda sa pretpostavljenim koeficijentima (1,5,10) iznosi 54,6 %, ocjena fluentnosti 24.4 %, dok je sveukupna ocjena prijevoda -21 % . Budući da su koeficijenti za velike i kritične pogreške vrlo veliki u pretpostavljenoj formuli, ocjene prijevoda također su izračunate sa manjim koeficijentima: 1 za male pogreške, 3 za velike i 5 za kritične. U tom slučaju ocjena fluentnosti iznosila je 54,6 %, ocjena točnosti 74,8 %, a sveukupna ocjena 29,4 %. Ocjene izračunane prema prilagođenoj formuli znatno bolje odgovaraju ocjenama fluentnosti i točnosti iz DQF-ova alata za evaluaciju kvalitete, koje iznose 2.5 za fluentnost i 3 za točnost. Razlog za vrlo niske ocjene s pretpostavljenim koeficijentima vjerojatno leži u tome što je Višedimenzionalna metrika izvorno namijenjena za označavanje gotovo savršenih (*near miss*) segmenata, pa ju je za svrhe ovakvih istraživanja potrebno prilagoditi. U sljedećim su tablicama pregledno prikazani rezultati istraživanja i ocjene.

Vrsta pogreške	Total	Σ	Σ krit. grešaka	Krit. greške	Σ vel. grešaka	Vel. greške	Σ mal. grešaka	Mal. greške
Sve kategorije	178	0	23	0	137	0	18	0
Točnost	56	0	18	0	33	0	5	0
Pogrešan prijevod	42	42	14	14	23	23	5	5
Izostavljeno	8	8	4	4	4	4	0	0
Neprevedeno	2	2	0	0	2	2	0	0
Dodano	4	4	0	0	4	4	0	0
Fluentnost	122	6	5	0	104	6	13	0
Gramatika	91	0	2	0	85	0	4	0
Oblik riječi	91	5	2	0	85	5	4	0
Vrsta riječi	13	13	1	1	11	11	1	1
Slaganje	70	70	0	0	67	67	3	3
Vrijeme/aspekt	3	3	1	1	2	2	0	0
Funkcijske riječi	6	0	2	0	4	0	0	0
Nepotrebno	1	1	0	0	1	1	0	0
Netočno	5	5	2	2	3	3	0	0
Pravopis	9	9	0	0	0	0	9	9
Tipografija	1	1	0	0	1	1	0	0
Redoslijed riječi	9	9	1	1	8	8	0	0

Tablica 4: Rezultati - psihosocijalna pomoć, Google Translate

	Predodređeni koeficijenti	Prilagođeni koeficijenti
Ocjena fluentnosti	24,4 %	54,6 %
Ocjena točnosti	54.6 %	74,8 %
Ukupna ocjena	-21 %	29,4 %

Tablica 5: Ocjene - psihosocijalna pomoć, Google Translate

U prijevodu alata Bing Translator pronađeno je sveukupno 199 pogrešaka, dakle veći broj pogrešaka nego u izlazu Googleova alata. Većina pogrešaka pripada kategoriji fluentnosti, gdje su pronađene 144 pogreške. To znači da su se u prijevodu pomoću alata Bing Translator nalazile čak 22 greške u fluentnosti više nego u izlazu alata Google Translate. Međutim, u usporedbi s drugim alatom, broj pogrešaka u točnosti gotovo je jednak: u izlazu alata Bing Translator pronađena je samo jedna pogreška manje. To pokazuje da je najveća razlika između Googleovog i Bingovog alata u tome što Google Translate proizvodi tekstove malo bolje fluentnosti. Kao i u prethodnom slučaju, većina pogrešaka u točnosti odnosi se na kategoriju pogrešnog prijevoda, dok većina grešaka u fluentnosti pripada kategoriji slaganja u rodu, broju i padežu. Iz perspektive težine pogrešaka, u ovom je izlazu pronađena 41 mala pogreška, 136 velikih i 22 kritične. Male pogreške, kojih ima preko dvostruko više nego u izlazu drugog alata, većinom uključuju greške u tipografiji i pravopisu (veliko i malo slovo).

Ocjena točnosti prijevoda u ovom alatu s pretpostavljenim koeficijentima iznosi 55,9 %, čime se odražava činjenica da izlaz alata Bing Translator sadrži manje pogrešaka u točnosti, tj. da izlaz alata Google Translate ima jednu kritičnu grešku u točnosti više. Ocjena fluentnosti iznosi 12,7 %, što je upola manje od ocjene fluentnosti za Google Translate. Sveukupna ocjena prijevoda pomoću alata Bing Translator iznosi -31,4%, što označava da je ovaj prijevod u cjelini lošiji od prijevoda pomoću drugog alata. Radi primjerenosti ovom konkretnom zadatku, izračunane su i ocjene s ugođenim koeficijentima . U tom slučaju, ocjena točnosti za Bing Translator iznosi 75,3 %, ocjena fluentnosti 46,6 %, a sveukupna ocjena 21,9 %. Ti nam rezultati pokazuju, međutim, da razlike u fluentnosti i sveukupnoj kvaliteti nisu toliko drastične kako se čini s pretpostavljenim koeficijentima, a taj rezultat mnogo bolje odražava sveukupni dojam anotatora i bolje odgovara ocjenama postignutoj u DQF alatima. Rezultati istraživanja i ocjene pregledno su prikazani u sljedećim tablicama.

Vrsta pogreške	Total	Σ	Σ krit. grešaka	Krit. greške	Σ vel. grešaka	Vel. greške	Σ mal. grešaka	Mal. greške
Sve kategorije	**199**	**0**	**22**	**0**	**136**	**0**	**41**	**0**
Točnost	55	0	17	0	27	0	11	0
Pogrešan prijevod	42	42	13	13	20	20	9	9
Izostavljeno	9	9	4	4	5	5	0	0
Neprevedeno	2	2	0	0	1	1	1	1
Dodano	2	2	0	0	1	1	1	1
Fluentnost	144	3	5	0	109	3	30	0
Gramatika	96	0	2	0	92	0	2	0
Oblik riječi	96	9	2	0	92	9	2	0
Vrsta riječi	13	13	1	1	12	12	1	1
Slaganje	73	73	1	1	70	70	2	2
Vrijeme/aspekt	1	1	0	0	1	1	0	0
Funkcijske riječi	6	0	0	0	6	0	0	0
Nedostaje	3	3	0	0	3	3	0	0
Netočno	3	3	0	0	3	3	0	0
Pravopis	10	10	0	0	1	1	10	10
Tipografija	11	11	0	0	2	2	9	9
Redoslijed riječi	18	18	3	3	6	6	9	9

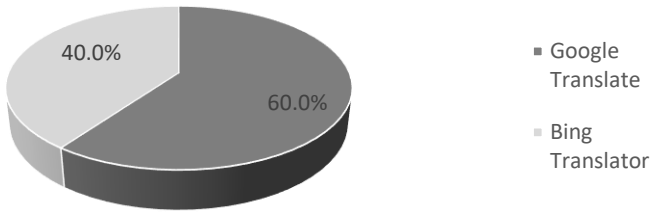
Tablica 6: Rezultati - psihosocijalna pomoć, Bing Translator

	Predodređeni koeficijenti	Prilagođeni koeficijenti
Ocjena fluentnosti	12,7 %	46,6 %
Ocjena točnosti	55,9 %	75,3 %
Ukupna ocjena	-31,4 %	21,9 %

Tablica 7: Ocjene - psihosocijalna pomoć, Bing Translator

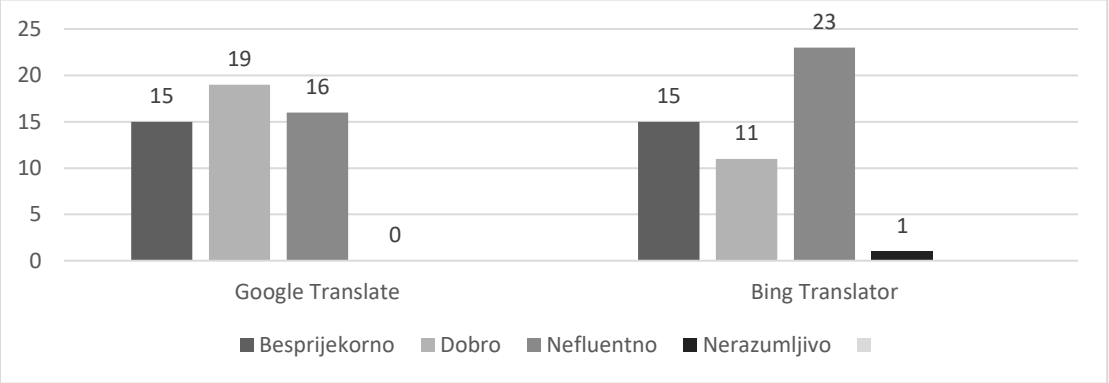
6.2 Rezultati istraživanja – Europska unija

Rezultati uspoređivanja izlaza dvaju alata na primjeru teksta koji pripada području Europske unije pokazuju da su svi evaluatori odabrali izlaz alata Google Translate u preko 58 % slučajeva. Međutim, sveukupna ocjena izlaza alata Bing Translator malo je bolja nego na primjeru prethodnog teksta. Sveukupni omjer prikazan je u sljedećem grafikonu.

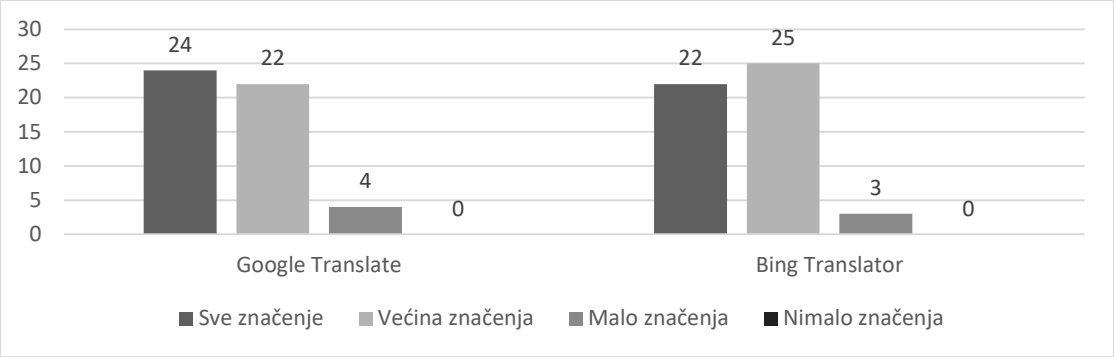


Grafikon 4: Preferirani izlaz alata - EU

U evaluaciji kvalitete fluentnost izlaza alata Google Translate ocjenjena je ocjenom 3 od 4, dok je fluentnost izlaza alata Bing Translator ocjenjena ocjenom 2. Iako je ocjena za Bingov alat jednaka kao i kod prethodnog teksta, Googleov alat zapravo je proizveo tekst bolje fluentnosti nego na primjeru prethodnog teksta: ocjena 3 označava da je tekst dobre fluentnosti. Alat Google Translate također je dobio ocjenu adekvatnosti 4, što znači da je značenje u potpunosti preneseno u prijevodu. Bing Translator ponovno je dobio ocjenu 3 za adekvatnost, po čemu se Google Translate pokazao boljim alatom po pitanju točnosti, a osim toga se pokazao i prikladnijim za prijevod tekstova iz područja Europske unije nego iz područja psihosocijalne pomoći. Distribucija fluentnosti i adekvatnosti za područje Europske unije pokazana je u sljedećim grafikonima.



Grafikon 5: Distribucija fluentnosti



Grafikon 6: Distribucija adekvatnosti

Rezultati prebrojavanja pogrešaka nažalost ponovno nisu pokazali neke značajne razlike među alatima. Google Translate ponovno ima malo više grešaka u točnosti, dok Bing Translator opet ima više grešaka u fluentnosti. No broj grešaka previše je sličan za oba alata da bi se mogli donositi ikakvi čvrsti zaključci. Međutim, moguće je usporediti sveukupan broj pogrešaka s brojem pogrešaka u prethodnom uzorku teksta slične duljine i primijetiti da se u prijevodu teksta iz drugog područja pojavljuje značajno manje grešaka nego u prijevodu teksta iz područja psihosocijalne pomoći.

	Točnost	**Fluentnost**	**Sveukupno**
Google	36	71	107
Bing	34	74	108

Tablica 8: Broj pogrešaka u tekstu o EU

Rezultati označavanja u alatu translate5 pokazali su da se u Googleovom prijevodu nalazi 123 grešaka, od kojih 38 pripada grani točnosti, a 85 grani fluentnosti. Najčešće su pogreške ponovno pogrešan prijevod i slaganje u rodu, broju i padežu. Međutim, u usporedbi s prethodnim tekstom tih grešaka ima osjetno manje. Dok pogrešnih prijevoda ima 34 (prethodni tekst sadrži 42), pogrešaka u slaganju riječi ima 42, što je značajno manje od 70, koliko ih se nalazi u prijevodu teksta o psihosocijalnoj pomoći. Osim toga, kritičnih grešaka ima samo 5, što je mnogo manje od 23 greške u prethodnom tekstu.

Ocjene ovog prijevoda također su bolje nego ocjene prijevoda prethodnog teksta. S pretpostavljenim koeficijentima ocjena točnosti je 76,1 %, ocjena fluentnosti 50,4 %, a sveukupna ocjena iznosi 26.5 %. S prilagođenim koeficijentima ocjene iznose 85,9 %, 69,8 % i 55,7 %. Te

ocjene pokazuju da je prijevod teksta o Europskoj uniji pomoću alata Google Translate u svakom pogledu osjetno bolji od prijevoda teksta o psihosocijalnoj pomoći.

Ocjene i rezultati prikazani su u sljedećim tablicama.

Vrsta pogreške	Total	Σ	Σ krit. grešaka	Krit. greške	Σ vel. grešaka	Vel. greške	Σ mal. grešaka	Mal. greške
Sve kategorije	**123**	**0**	**5**	**0**	**101**	**0**	**17**	**0**
Točnost	38	0	4	0	28	0	6	0
Pogrešan prijevod	24	24	4	4	15	15	5	5
Izostavljeno	7	7	0	0	7	7	0	0
Dodano	7	7	0	0	6	6	1	1
Fluentnost	85	2	1	0	73	1	11	1
Gramatika	61	0	1	0	58	0	2	0
Oblik riječi	61	14	1	0	58	14	2	0
Vrsta riječi	4	4	0	0	4	4	0	0
Slaganje	42	42	1	1	39	39	2	2
Vrijeme/aspekt	1	1	0	0	1	1	0	0
Funkcijske riječi	9	0	0	0	8	0	1	0
Nepotrebno	1	1	0	0	1	1	0	0
Nedostaje	3	3	0	0	3	3		
Netočno	5	5	0	0	4	4	1	1
Pravopis	4	4	0	0	0	0	4	4
Tipografija	3	3	0	0	0	0	3	3
Redoslijed riječi	6	6	0	0	6	6	0	0

Tablica 9: Rezultati - EU, Google Translate

	Predodređeni koeficijenti	Prilagođeni koeficijenti
Ocjena fluentnosti	50, 4 %	69, 8 %
Ocjena točnosti	76,1 %	85, 9 %
Ukupna ocjena	26, 5 %	55, 7 %

Tablica 10: Ocjene - EU, Google Translate

U prijevodu Bingova alata pronađeno je sveukupno 128 pogrešaka, slično kao i u prijevodu alata Google Translate. Grani točnosti pripadaju 43 pogreške, a grani fluentnosti 85 pogrešaka. Međutim, iako je broj pogrešaka u fluentnosti jednak kao i kod prethodnog alata, prijevod alata Bing Translate sadrži veći broj velikih grešaka u fluentnosti (78), što njegovu ocjenu fluentnosti čini nižom. S pretpostavljenim koeficijentima, ocjena točnosti za Bingov prijevod iznosi 69 %, ocjena fluentnosti 44,1 %, a sveukupna ocjena je 13,1 %. S prilagođenim koeficijentima ocjene su 82 %, 66,3 % i 48,2 %. Iz toga možemo zaključiti da je Google Translate postigao bolje rezultate na svim razinama, iako razlika nije velika. Svi su podaci prikazani u sljedećim tablicama.

Vrsta pogreške	**Total**	**Σ**	**Σ krit. grešaka**	**Krit. greške**	**Σ vel. grešaka**	**Vel. greške**	**Σ mal. grešaka**	**Mal. greške**
Sve kategorije	**128**	**0**	**7**	**0**	**110**	**0**	**11**	**0**
Točnost	43	0	6	0	20	20	4	4
Pogrešan prijevod	30	30	6	6	20	20	4	4
Izostavljeno	10	10	0	0	10	10	0	0
Dodano	3	3	0	0	2	2	1	1
Fluentnost	85	1	1	0	78	1	6	0
Gramatika	64	0	1	0	62	0	1	0
Oblik riječi	64	11	1	0	62	11	1	0
Vrsta riječi	5	5	0	0	5	5	0	0
Slaganje	45	45	1	1	43	43	1	1
Vrijeme/aspekt	3	3	0	0	3	3	0	0
Funkcijske riječi	4	0	0	0	4	0	0	0
Nepotrebno	1	1	0	0	1	1	0	0
Netočno	3	3	0	0	3	3	0	0
Pravopis	3	3	0	0	0	0	3	3
Tipografija	3	3	0	0	1	1	2	2
Redoslijed riječi	10	10	0	0	10	10	0	0

Tablica 11: Rezultati - EU, Bing Translator

	Predodređeni koeficijenti	Prilagođeni koeficijenti
Ocjena fluentnosti	44,1 %	66,3 %
Ocjena točnosti	69 %	82 %
Ukupna ocjena	13,1 %	48,2 %

Tablica 12: Ocjene - EU, Bing Translator

7. Zaključak

Istraživanje je pokazalo da alat Google Translate dosljedno daje bolje prijevode u pogledu fluentnosti, dok je alat Bing Translate proizveo zanemarivo bolju točnost izlaza na primjeru teksta o psihosocijalnoj pomoći. Bez obzira na vrstu teksta i alate, najčešće su pogreške u prijevodu vezane s pogrešnim prijevodom i slaganjem riječi u rodu, broju i padežu. Kada je riječ o vrsti teksta, oba su alata postigla osjetno bolje rezultate na primjeru teksta o Europskoj uniji, što se lako da objasniti činjenicom da se na internetu nalazi velik broj tekstova te tematike. Zbog politike EU-a prema kojoj sve informacije moraju biti dostupne svim građanima EU-a na njihovu jeziku, većina tih tekstova je prevedeno, zbog čega su statistički alati za strojno prevođenje naučili mnogo više tekstova iz tog područja.

Ovo istraživanje također otkriva nekoliko činjenica o alatima koji su se koristili. DQF-ov alat za uspoređivanje segmenata pokazao se dostatnim u slučajevima kada je potrebna brza metoda ljudske evaluacije za koju je lako pronaći evaluatore. Međutim, ta metoda samo daje samo uvid u to koji je alat za strojno prevođenje „bolji“, a ne omogućuje dublju analizu pomoću većeg broja metrika. Detaljniju analizu pokušava pružiti DQF-ov alat za evaluaciju kvalitete, ali u ovom se istraživanju pokazalo da je korisno samo njegovo ocjenjivanje fluentnosti i adekvatnosti. Funkcionalnost prebrojavanja tipoloških pogrešaka nije dala korisne rezultate zbog toga što je broj pogrešaka uvijek bio sličan za oba alata. No to ne znači da je ova funkcionalnost beskorisna. Prebrojavanje pogrešaka ionako nije zamišljeno kao metoda uspoređivanja dvaju sustava, već samo kao uvid u količinu grešaka u jednom prijevodu. Osim toga, moguće je da bi se na većem uzorku također pokazale veće razlike.

Označavanje tekstova pomoću Višedimenzionalne metrike u teoriji pruža najdublju razinu analize. Iako je ta metoda omogućila uvid u detaljnije vrste pogrešaka i njihovu težinu, ispostavilo se da su najčešće vrste pogrešaka u svim prijevodima jednake. Osim toga, pretpostavljeni koeficijenti u formuli za izračunavanje završne ocjene pokazali su se neprikladnima za

označavanje tekstova s velikom količinom grešaka, zbog čega je bilo potrebno koristiti se prilagođenom formulom. Ocjene koje su dobivene na taj način pokazale su realne rezultate evaluacije fluentnosti, točnosti i sveukupne kvalitete. Međutim, budući da ova metoda evaluacije iziskuje veliku količinu vremena, truda, znanja i vještine, upitno je je li isplativa za ovakvu vrstu zadatka.

Općenito, metode ljudske evaluacije učinkovit su način za vrednovanje strojnih prijevoda u slučajevima kada vrijeme i novac nisu problem. Te metode omogućuju vrednovanje različitih parametara prijevoda ovisno o svrsi istraživanja te pružaju uvid u značajke strojnih prijevoda koje automatske metrike ne mogu prikazati. U sklopu Dinamičkog okvira i Višedimenzionalnih metrika pruženi su alati kojima se može obavljati većina zadataka svojstvenih ljudskoj evaluaciji: od jednostavnog rangiranja do složene analize pogrešaka. U skladu s potrebama ti se alati mogu nadopunjavati te se njihovi rezultati mogu nadovezivati. Međutim, kod korištenja kompleksnijih metrika također je vrlo bitno od početka jasno definirati ciljeve istraživanja kako se ne bi koristile redundantne metrike i alati.

8. Literatura

1. Adequacy/Fluency Guidelines. Svibanj 2013. *Quality Evaluation using Adequacy and/or Fluency Approaches*. https://www.taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluency-guidelines.6.9.2016.

2. Banerjee, S.; Lavie, A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements. Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Comupational Linguistics. Michigan, 2005.

3. Brkić, Marija; Seljan, Sanja; Matetić, Maja. Machine Translation Evaluation for Croatian-English and English-Croatian Language Pairs. // Proceedings fo the 8th International NLPCS Workshop: Human-Machine Interaction in Translation / Sharp, Bernardette ; Zock, Michael ; Carl, Michael ; Jakobsen, Arnt Lykke (ur.). Copenhagen : Copenhagen Business School, 2011. 93-104

4. Brkić, Marija; Seljan, Sanja; Vičić, Tomislav. Automatic and Human Evaluation on English-Croatian Legislative Test Set. // Lecture Notes in Computer Science - LNCS. 7816 (2013) , 1; 311-317

5. Brkić, Marija; Vičić, Tomislav; Seljan, Sanja. Evaluation of the Statistical Machine Translation Service for Croatian-English. // 2nd international conference: The future of information sciences (INFuture 2009) : Digital resources and knowledge sharing : proceedings / Stančić, Hrvoje ; Seljan, Sanja ; Bawden, David ; Lasić-Lazić, Jadranka ; Slavić, Aida (ur.). Zagreb : Department of Information Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2009. 319-332

6. Burchardt, A.; Lommel A. Practical Guidelines for the Use of MQM in Scientific Research on Translation Quality. 2014.

7. Creating a translation quality score with MQM. http://www.qt21.eu/launchpad/node/1332. 6.9.2016.

8. Creating and annotating projects | Q Launchpad. http://www.qt21.eu/launchpad/node/1342#. 6.9.2016.

9. Data formats | QTLaunchpad. http://www.qt21.eu/launchpad/node/1345#. 6.9.2016.

10. Denkowski, M.; Lavie, A. Choosing the Right Evaluation for Machine Translation: an Examination of Annotator and Automatic Metric Performance on Human Judgement Tasks. // Proceedings of the Ninth Biennial Conference of the Association for Machine Translation in the Americas. Denver, Colorado. 2010.

11. DePalma, D. A. LISA Shuts Down Operations. 28. 2. 2011. http://www.commonsenseadvisory.com/Default.aspx?Contenttype=ArticleDetAD&tabID=63&Aid=1357&moduleId=390. 6.9.2016.

12. Gallafent, A. Machine translation for the military. http://www.theworld.org/2011/04/machine-translation-military/. 18.9.2016.

13. Hutchins, J. The development and use of machine translation systems and computer-based translation tools. // International Symposium on Machine Translation and Computer Language Information Processing. Beijing: China, 1999. Str 26-28.

14. Jackson, W. Air Force wants to build a universal translator. https://gcn.com/articles/2003/09/09/air-force-wants-to-build-a-universal-translator.aspx. 18.9.2016.

15. Jurafsky, D.; Martin H. J. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. New Jersey: Pearson education, 2009.

16. Koehn, P.; Monz, C. Manual and Automatic Evaluation of Machine Translation between European Languages. // Proceedings of the Workshop on Statistical Machine Translation. New York City : Association for Computational Linguistics, 2006. Str. 102-121.

17. Logging in and annotating | QTLaunchpad. http://www.qt21.eu/launchpad/node/1339. 6.9.2016.

18. Lommel, A. Multidimensional Quality Metrics (MQM): A New Framework for Translation Quality Assessment. Prezentacija. 2014.

19. Lommel, A. Multidimensional Quality Metrics. Meta Forum 2013. Prezentacija. 2013.

20. Lommel, A.; Uszkoreit, H.; Burchardt, A. Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. Revista Tradumàtica: tecnologies de la traducció. 2014. Str. 455-462.

21. Machine Translation Service. http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm. 18.9.2016.

22. Multidimensional Quality Metrics (MQM) Definition. http://www.qt21.eu/mqm-definition/definition-2015-12-30.html. 6.9.2016.

23. Multidimensional Quality Metrics (MQM) Issue Types. http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html. 6.9.2016.

24. Multilingual Online Translation. http://www.molto-project.eu/. 19.9.2016.

25. Pospelova, O.; Rowda, J. Human Evaluation of Machine Translation. 26.6.2016. http://www.ebaytechblog.com/2016/06/26/human-evaluation-of-machine-translation/. 6.9.2016.

26. Resources | QTLaunchpad. http://www.qt21.eu/launchpad/content/resources. 6.9.2016.

27. Second Machine Translation Marathon. Bilješke s predavanja. Njemačka : Berlin, 2008.

28. Seljan, S.; Vičić, T.; Brkić, M. BLEU Evaluation of Machine-Translated English-Croatian Legislation. // Proceedings of the Eighth International Conference on Language Resources and Evaluation. Istanbul, Turkey: European Language Resources Association, 2012.

29. Seljan, Sanja; Brkić, Marija; Kučiš, Vlasta. Evaluation of Free Online Machine Translations for Croatian-English and English-Croatian Language Pairs. // INFuture2011: The Future of Information Sciences - Information Sciences and e-Society / Billenness, Clive ; Hemera, Annette ; Mateljan, Vladimir ; Banek Zorica, Mihaela ; Stančić, Hrvoje ; Seljan, Sanja (ur.). Zagreb : Department of Information Sciences, 2011.. Str. 331-344

30. Seljan, Sanja; Dunđer, Ivan. Machine Translation and Automatic Evaluation of English/Russian-Croatian // Proceedings of the International Conference "Corpus Linguistics - 2015". St. Petersburg, Rusija : St. Petersburg State University, 2015. 72-79

31. Setting up a translate5 project | QTLaunchpad. http://www.qt21.eu/launchpad/node/1336. 6.9.2016.

32. Što je to prijevod i prevođenje,a što je strojno prevođenje?. http://www.prevoditelj.com/prijevod-prevodjenje-i-strojno_prevodjenje/. 18.9.2016.

33. TAUS - Mission. https://www.taus.net/mission. 6.9.2016.

34. TAUS launches Dynamic Quality Evaluation Framework. 12.6.2016. https://www.taus.net/think-tank/news/press-release/taus-launches-dynamic-quality-evaluation-framework. 6.9.2016.

35. The measurement of observer agreement for categorical data. // Biometrics / Landis, J. R.; Koch, G. G. Washington DC : The International Biometric Society. 1977. Str. 159-174.

36. User management | QTLaunchpad. http://www.qt21.eu/launchpad/node/1338. 6.9.2016.

37. Vilar, D.; Leusch, H. N.; Banchs, R. Human evaluation of machine translation through binary system comparisons. // *ACL2007 SMT Workshop*. 2007.

8.1. Popis slika

8.2. Popis tablica

8.3. Popis grafikona

8.4.. Popis priloga

MQM Evaluation Scorecard Current score: 100% Logged in as: • logout

Scorecard | Project info | Project specifications | Reports | Training and help

✔ All changes saved

	Source: 1 of 211	Target: 1 of 211
73		CARICOM Regional Code of Practice for Food Hygiene
74		CODEX Alimentarius Commision,CAC/RCP 44 Code of Practice for packaging and transport of tropical fresh fruits and vegetables
75	2 Terms and definitions	3 Termes et définitions
76	For the purposes of this standard the following terms and definitions shall apply.	Les termes et définitions suivants sont applicables pour cette norme.

Notes

Save Note

Navigation

Go to seg: seg. num. Go

Accuracy: Accuracy | Omission | Addition | Mistranslation | Untranslated

Fluency: Fluency | Typography | Spelling | Grammar | Word order | Word form

Part of speech | Agreement | Tense/mood/aspect | Function words | Extraneous | Missing

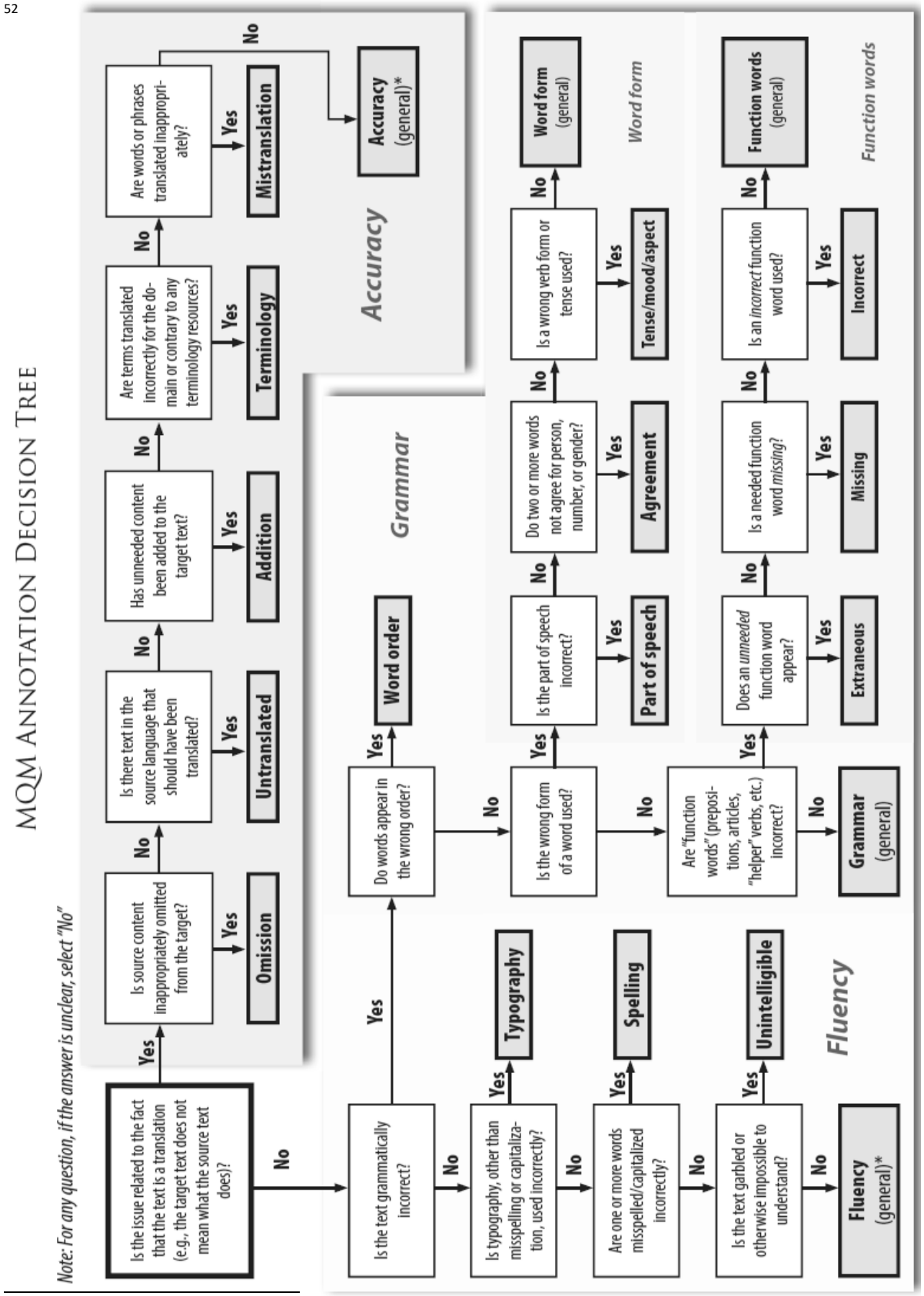
Incorrect | Unintelligible

Segment list and editor

Editor modes | Hide tags | Short tag view | Full tag view | Reset sorting / filtering (50) | watchlist | QM Subsegment Statistics

Nr.	Autostatus	source	Google Translate	Google Translate	Bing Translator	Bing Translator	Comments
1		Action Sheet Nr. 1: MHPSS Core Principles	Akcija list br. 1: MHPSS Osnovna načela	Akcija [14] list [14] br. 1: [12] MHPSS [12] Osnovna načela	Akcija list Nr. 1: MHPSS osnovna načela	Akcija [14] list [14] [4] Nr [4] . 1: [12] MHPSS [12] osnovna načela	
2		Area	područje	[7] p [7] odručje	Područje	Područje	
3		All event types, all target groups, all phases	Svi tipovi događaja, sve ciljne skupine, u svim fazama	Svi tipovi događaja, sve ciljne skupine, u svim fazama	Svih vrsta događaja, sve ciljane skupine, sve faze	[12] Svih vrsta [12] događaja, sve ciljane skupine, sve faze	
4		MHPSS core principles in both IASC and NATO TENTS guidelines	temeljna načela MHPSS u oba IASC-a i NATO šatora smjernicama	temeljna načela [12] MHPSS [12] u [2] oba [2] [12] IASC-a [12] i NATO [2] šatora [2] smjernicama	MHPSS osnovnih načela u IASC i NATO šatora smjernice	[12] MHPSS [12] [12] osnovnih načela [12] u IASC i NATO [2] šatora [2] [12] smjernice [12]	
5		Principle 1: Ensure human rights and equity	1. načelo: Osigurati ljudska prava i jednakost	1. načelo: Osigurati ljudska prava i jednakost	Princip 1: Osigurati ljudska prava i jednakost	[14] Princip [14] [8] 1 [8] : Osigurati ljudska prava i jednakost	
6		Humanitarian actors should promote the human rights of all affected persons and protect individuals and groups who are at heightened risk of human rights violations and at the same time ensure participation.	Humanitarni glumci trebali promicati ljudska prava svih pogođenih osoba i štite pojedince i skupine koje su pod povećanim rizikom od kršenja ljudskih prava, au isto vrijeme osigurati sudjelovanje.	Humanitarni [2] glumci [2] [10] trebali [10] promicati ljudska prava svih pogođenih osoba i [10] štite [10] pojedince i skupine [12] koje [12] su pod povećanim rizikom od kršenja ljudskih prava, [7] au [7] isto vrijeme osigurati sudjelovanje.	Humanitarni akteri trebaju ljudskih prava za sve osobe, zaštite pojedinaca i skupinakoje su na povećanje rizika od kršenja ljudskih prava i u isto vrijeme osigurati sudjelovanje.	Humanitarni akteri trebaju [3] 3 [3] [12] ljudskih [12] prava za sve osobe, [11] zaštite [11] [12] pojedinaca [12] i [12] skupinakoje [12] [7] [7] su [18] na [18] [11] povećanje [11] [12] rizika [12] od kršenja ljudskih prava i u isto vrijeme osigurati sudjelovanje.	
7		Principle 2: Do no harm	2. načelo: ne naškoditi	2. načelo: ne naškoditi	Princip 2: Ne nauditi	[14] Princip [14] [8] 2 [8] : Ne nauditi	
8		Humanitarian aid is an important means of helping people affected by emergencies, but aid can also cause unintentional harm.	Humanitarna pomoć je važan način pomaže oboljelima od hitne slučajeve, ali potpora također može uzrokovati nenamjerne štete.	Humanitarna pomoć je važan način [11] pomaže [11] oboljelima od [12] hitne slučajeva [12] , ali potpora također može	Humanitarna pomoć je važno sredstvo za pomaganje ljudima pogođenim hitnim slučajevima, ali pomoć također mogu	Humanitarna pomoć je važno sredstvo za pomaganje ljudima pogođenim hitnim slučajevima, ali pomoć također [12] mogu	

Prilog 2: Sučelje alata translate5



Prilog 3: Stablo odlučivanja za MQM