

Evaluacija strojnih prijevoda s njemačkoga na hrvatski jezik

Tonković, Kristina

Master's thesis / Diplomski rad

2019

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka / Sveučilište u Rijeci**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:195:204770>

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

Download date / Datum preuzimanja: **2025-03-14**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Informatics and Digital Technologies - INFORI Repository](#)



Sveučilište u Rijeci – Odjel za informatiku

Diplomski jednopredmetni studij informatike: nastavnički smjer

Kristina Tonković

Evalvacija strojnih prijevoda s njemačkoga na hrvatski jezik

Diplomski rad

Mentor: doc. dr. sc. Marija Brkić Bakarić

Rijeka, srpanj 2019.

Rijeka, 7.6.2019.

Zadatak za diplomski rad

Pristupnik: Kristina Tonković

Naziv diplomskog rada: Evaluacija strojnih prijevoda s njemačkoga na hrvatski jezik

Naziv diplomskog rada na eng. jeziku: Evaluation of machine translations from German to Croatian

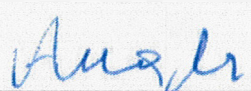
Sadržaj zadatka:

Strojni prijevodi, iako postaju sve razvijeniji i kvalitetniji, još uvijek mogu rezultirati pogreškama koje se mogu procijeniti i ispraviti samo ljudskom evaluacijom. Ponekad, primjerice, prijevodi mogu biti gramatički posve točni, međutim, mogu sadržavati semantičke pogreške. Zadatak diplomskog rada obuhvaća automatsku evaluaciju i analizu pogrešaka strojnih prijevoda triju tekstova iz različitih domena s njemačkoga na hrvatski jezik. Analizom će se obuhvatiti strojni prevoditelji Google Translate i Bing Microsoft Translator, a u automatskoj evaluaciji koristit će se metrike BLEU i chrF++. Pearsonovim koeficijentom korelacije utvrdit će se povezanost rezultata odabranih automatskih metrika i analize pogrešaka izvršene u alatu TREAT, a u skladu s MQM okvirom.

Mentor:
doc. dr. sc. Marija Brkić Bakarić

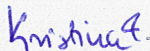


Voditeljica za diplomske radove:
izv. prof. dr. sc. Ana Meštrović



Komentor:

Zadatak preuzet: 10.6.2019.


(potpis pristupnika)

Sažetak

Razvoj tehnologije, informatizacija i modernizacija društva omogućili su, između ostalog, olakšavanje poslova prevodjenja tekstova s jednoga jezika na drugi. Razni softverski programi, a također i online prevoditelji razvijeni su do visoke razine. Međutim, postavlja se pitanje koliko se zapravo korisnik takvih prevoditelja može osloniti na njih te je li ipak još uvijek potreban i ljudski faktor – osoba koja će takve prijevode kontrolirati i prepravljati.

Diplomski rad donosi izračun BLEU i chrF++ računalne metrike za prijevode koji pripadaju trima vrstama teksta – recept, upute i vijesti. Tekstovi su prevedeni pomoću Bing i Google online prevoditelja s njemačkoga na hrvatski jezik, a uspoređivani su s ljudskim prijevodom tekstova.

Također je nad prijevodima izvedena i ljudska evaluacija pomoću MQM metrike, a osim toga izvedeni su i izračuni Pearsonovog koeficijenta korelacije za varijable koje se odnose na rezultate BLEU i chrF++ metrike na razini rečenice te broja pogrešaka koje su otkrivene ljudskom evaluacijom, također na razini rečenice.

Rezultati su pokazali da su strojni prevoditelji zaista na visokoj razini, međutim još uvijek se javljaju greške koje su ponekad ljudskom evaluatoru posve besmislene i neočekivane.

Pogreške se javljaju u području sintakse, kao i u području semantike, zbog čega se korisnik strojnih prevoditelja još uvijek ne može u potpunosti „osloniti“ na njih.

Cljučne riječi: strojni prijevod, evaluacija, BLEU metrika, chrF++ metrika, MQM metrika, Bing Microsoft Translator, Google Translate, Pearsonov koeficijent korelacije

Sadržaj

| | |
|---|----|
| 1. Uvod | 1 |
| 2. BLEU metrika | 2 |
| 2.1. Računanje BLEU metrike..... | 2 |
| 2.1.1. Kumulativni rezultat BLEU metrike..... | 3 |
| 3. BLEU metrika – tekstovi za analizu..... | 4 |
| 3.1. Recept..... | 4 |
| 3.1.1. Corpus BLEU rezultati – recept | 5 |
| 3.2. Upute | 6 |
| 3.2.1. Corpus BLEU rezultati – upute | 6 |
| 3.3. Vijesti | 8 |
| 3.3.1. Corpus BLEU rezultati – vijesti | 8 |
| 3.4. BLEU rezultati - zaključak | 9 |
| 4. Character n-gram F-score | 11 |
| 4.1. chrF++ | 11 |
| 4.2. ChrF++ - rezultati..... | 12 |
| 4.2.1. Rezultati – Tablica..... | 13 |
| 5. Ljudska evaluacija na temelju MQM-a | 15 |
| 5.1. MQM metrika | 15 |
| 5.2. Rezultati ljudske evaluacije | 17 |
| 5.2.1. Ljudska evaluacija – tablica | 18 |
| 5.3. Usporedba pogrešaka strojnih prijevoda | 20 |
| 5.3.1. Vrsta teksta – recept..... | 20 |
| 5.3.2. Vrsta teksta – upute..... | 22 |
| 5.3.3. Vrsta teksta – vijesti..... | 24 |
| 6. Pearsonov koeficijent korelacije | 27 |
| 6.1. Pearsonov koeficijent korelacije: BLEU metrika | 28 |

| | |
|--|----|
| 6.1.1. Rezultati: Vrsta teksta – Recept | 28 |
| 6.1.2. Rezultati: Vrsta teksta – Upute..... | 30 |
| 6.1.3. Rezultati: Vrsta teksta – Vijesti..... | 32 |
| 6.2. Pearsonov koeficijent korelacije: chrF++ metrika | 35 |
| 6.2.1. Rezultati: Vrsta teksta – Recept | 35 |
| 6.2.2. Rezultati: Vrsta teksta – Upute..... | 37 |
| 6.2.3. Rezultati: Vrsta teksta – Vijesti..... | 39 |
| 6.3. Rezultati: Accuracy i Mistranslation – BLEU metrika | 41 |
| 6.3.1. Accuracy i Mistranslation – tekst „Recept“..... | 41 |
| 6.3.2.. Accuracy i Mistranslation – tekst „Upute“ | 43 |
| 6.3.3. Accuracy i Mistranslation – tekst „Vijesti“ | 45 |
| 6.4. Accuracy i Mistranslation – chrF++ metrika..... | 47 |
| 6.4.1. Accuracy i Mistranslation – tekst „Recept“..... | 48 |
| 6.4.2. Accuracy i Mistranslation – tekst „Upute“ | 50 |
| 6.4.3. Accuracy i Mistranslation – tekst „Vijesti“ | 52 |
| 6.4.4. Rezultati: Accuracy i Mistranslation – tablica..... | 54 |
| 6.5. Pearsonov koeficijent korelacije –diskusija | 55 |
| 7. Zaključak..... | 56 |
| 8. Literatura..... | 57 |
| POPIS SLIKA..... | 58 |
| POPIS TABLICA..... | 60 |

1. Uvod

Tema ovoga diplomskog rada obuhvaća projektni zadatak pod nazivom „Evaluacija strojnih prijevoda s njemačkoga na hrvatski jezik“ kojim se pokušalo utvrditi da strojno prevođenje, iako sve naprednije i kvalitetnije, još uvijek rezultira pogreškama koje se mogu procijeniti i ispraviti samo ljudskom evaluacijom.

Za projektni zadatak korišteni su programi Spyder, alat TREAT, BLEU i chrF++ metrika, MQM metrika, Pearsonov koeficijent korelacije, a strojni prevoditelji, čiji se prijevod evaluirao, jesu Google Translate i Bing Microsoft Translator.

Za analizu prijevoda korištene su tri različite vrste tekstova – recept, upute za Samsung mobilne uređaje te medijski članak tj. vijesti.

Prikazani su rezultati BLEU metrike za prijevode s njemačkoga na hrvatski jezik, dobiveni pomoću oba prethodno navedena online prevoditelja. Potom su prikazani rezultati chrF++ metrike, a nakon toga, u posebnom poglavlju, provedena je ljudska evaluacija prijevoda pomoću MQM metrike i definiranih smjernica za njeno korištenje. Pored toga, prikazane su neke od konkretnih pogrešaka prijevoda označene u alatu TREAT.

Također je izveden i izračun Pearsonovog koeficijenta korelacije za varijable koje uključuju rezultate dobivene računalnim metrikama BLEU i chrF++ na razini rečenice te broj pogrešaka prijevoda, također na razini rečenice.

Na kraju rada dan je zaključak teme, kao i osobno mišljenje o kvaliteti Google i Bing online prevoditelja.

2. BLEU metrika

BLEU metrika - *Bilingual Evaluation Understudy* – je metrika koja nam daje rezultat usporedbe strojnog prijevoda teksta i jednog ili više referentnih tekstova.¹

Ova metrika je danas jedna od najkorištenijih, a zamisao metrike je korištenje težinskog prosjeka uparenih fraza iz strojnog i referentnog prijevoda. Formula BLEU metrike jest

$$BLEU = kazna \times \exp \sum_{i=1}^n \lambda_i \log preciznost_i ,$$

a kazna se računa prema sljedećoj formuli:

$$e^{1 - \frac{\text{duljinaReferentnogPrijevoda}}{\text{duljinaStrojnogPrijevoda}}} .$$

Metrika je brza i lako razumljiva, savršen rezultat podudaranja tekstova pomoću ove metrike iznosi 1.0, a potpuno nepodudaranje daje rezultat 0.0.

BLEU metrika radi na način da se broje podudarajući n-grami u strojnom prijevodu s n-gramima u referentnom tekstu, gdje je 1-gram svaki „token“, odnosno svaka pojedina riječ, a 2-gram se odnosi na svaki par riječi, itd.

Savršeni rezultat, tj. savršeno podudaranje zapravo u praksi nije posve moguće, budući da bi to značilo da je strojni prijevod posve identičan referentnom (najčešće ljudskom) prijevodu.

„Osnovne zamjerke BLEU metrici su to što zanemaruje riječi kod kojih nije uporabljen pravilni oblik i uporabu sinonima. Nadalje, relativni značaj riječi ne uzima se u obzir. Primjerice, izostanak rječice „ne“ uvelike mijenja značenje rečenice, dok interpunkcija nije od presudne važnosti.“ (Brkić Bakarić 2018: 30)

2.1. Računanje BLEU metrike

Računanje rezultata BLEU metrike omogućuje NLTK Python biblioteka, tj. *Natural Language Toolkit* biblioteka, koja je također korištena i pri računanju rezultata uspoređivanih tekstova ovoga diplomskog rada.

¹ Izvor: Papineni, K., Roukos, S., Ward, T.,Zhu, W.(2002) *BLEU: a Method for Automatic Evaluation of Machine Translation*.

Biblioteka nudi `sentence_bleu()` funkciju koja daje rezultat usporedbe na razini rečenica, dok funkcija `corpus_bleu()` daje rezultat usporedbe na razini odlomka ili cijeloga dokumenta.

Primjer funkcije `sentence_bleu()`:

```
1 from nltk.translate.bleu_score import sentence_bleu
2 reference = [['ovo', 'je', 'prvi', 'primjer']]
3 candidate = ['ovo', 'je', 'prvi', 'primjer']
4 score = sentence_bleu(reference, candidate)
5 print(score)
6
```

Slika 1 - Primjer funkcije `sentence_bleu()`

Navedeni kod nam kao rezultat daje 1.0.

Primjer funkcije `corpus_bleu()`:

```
1 from nltk.translate.bleu_score import corpus_bleu
2 references = [['ReferentniTekst.txt']]
3 candidates = ['StrojniPrijevod.txt']
4 score = corpus_bleu(references, candidates)
5 print(score)
```

Slika 2 - Primjer funkcije `corpus_bleu()`

2.1.1. Kumulativni rezultat BLEU metrike

Kumulativni rezultat odnosi se na izračun individualnih n-gram rezultata u redoslijedu od 1 do n te na izračun njihove geometrijske sredine.

Prethodno navedene funkcije koriste se za računanje kumulativnog 4-gram rezultata, koji se također naziva i BLEU-4. Težine za 4-gram iznose 0.25, za 3-gram 0.33, za 2-gram 0.5, a za 1-gram težina iznosi 1. Primjerice:

```
1 # cumulative BLEU scores
2 from nltk.translate.bleu_score import sentence_bleu
3 reference = [['this', 'is', 'small', 'test']]
4 candidate = ['this', 'is', 'a', 'test']
5 print('Cumulative 1-gram: %f' % sentence_bleu(reference, candidate, weights=(1, 0, 0, 0)))
6 print('Cumulative 2-gram: %f' % sentence_bleu(reference, candidate, weights=(0.5, 0.5, 0, 0)))
7 print('Cumulative 3-gram: %f' % sentence_bleu(reference, candidate, weights=(0.33, 0.33, 0.33, 0)))
8 print('Cumulative 4-gram: %f' % sentence_bleu(reference, candidate, weights=(0.25, 0.25, 0.25, 0.25)))2
```

Slika 3 - izračun kumulativnog rezultata

Rezultati iznose redom: 0.75, 0.5, 0.0 i 0.0.

² Primjer preuzet sa <https://machinelearningmastery.com/calculate-bleu-score-for-text-python/> (8.7.2019.)

3. BLEU metrika – tekstovi za analizu

Za analizu, odnosno izračun BLEU metrike korištene su tri vrste tekstova koji obuhvaćaju područje recepata, uputa za Samsung mobilne uređaje te područje online članaka, točnije vijesti. Tekstovi su prethodno korišteni na kolegiju Jezične vježbe V i Jezične vježbe VI na Odsjeku za germanistiku Filozofskoga fakulteta u Rijeci, a pojedini dijelovi su spojeni iz više izvora, odnosno stvoreni su kombinacijom primjerice dva teksta (npr. dva recepta).

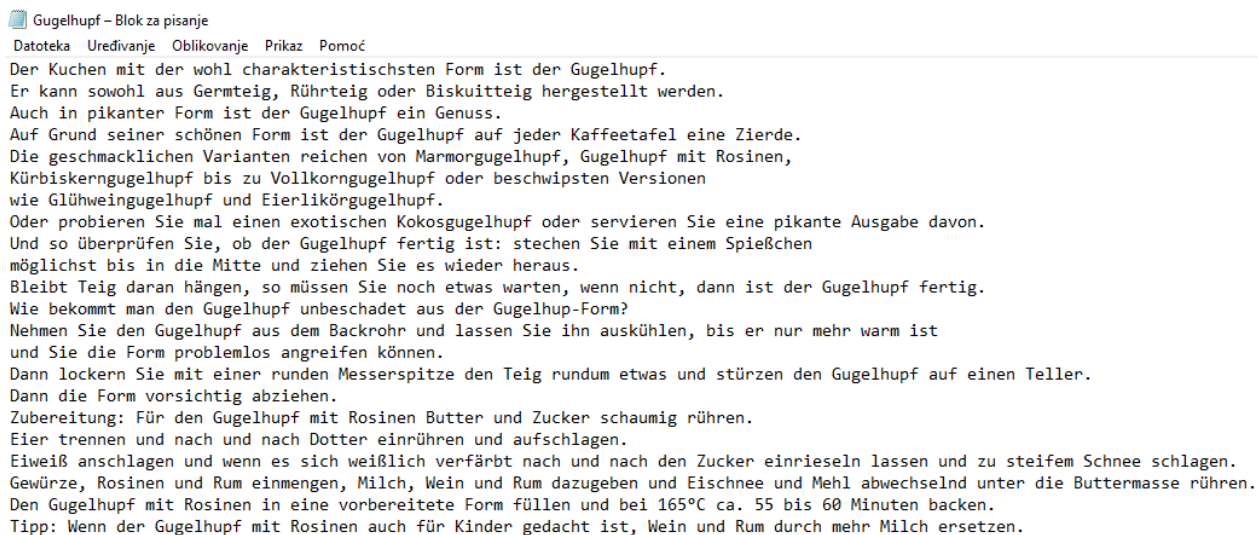
Tekstovi su prevedeni s njemačkoga na hrvatski jezik pomoću dva online prevoditelja: Google Translate i Bing Microsoft Translator.

Sljedeća potpoglavlja daju prikaz tekstova, a također i dobivene rezultate usporedbe prijevoda dobivenog pomoću svakog prethodno navedenog online prevoditelja s referentnim, odnosno ljudskim prijevodom. Korištena je naredba *corpus_bleu* koja daje BLEU-4 rezultate.

Za izračun je korišteno Spyder okruženje koje koristi programski jezik Python.

3.1. Recept

Obrađen je tekst pod nazivom „Gugelhupf“ – „Kuglof“.



Slika 4 - Prikaz teksta "Gugelhupf"

Slijede rezultati usporedbe prijevoda online prevoditelja s referentnim (ljudskim) prijevodima.

3.1.1. Corpus BLEU rezultati – recept

Dan je prikaz rezultata dobivenih korištenjem `corpus_bleu()` funkcije.

Mogući rezultati variraju od 0 do 1.

Bing Microsoft Translator

```
2 from nltk.translate.bleu_score import corpus_bleu
3 references = [['GugelhupfLjudski.txt']]
4 candidates = ['GugelhupfBing.txt']
5 score = corpus_bleu(references, candidates)
6 print(score)
7
```

Slika 5 - Prikaz koda za `corpus_bleu()` – recept - Bing

```
In [5]: runfile('C:/Users/Kristina/.spyder-py3/temp.py', wdir='C:/Users/
Kristina/.spyder-py3')
0.538118676268092
```

Slika 6 - Rezultat `corpus_bleu()` - recept - Bing

Google Translate

```
2 from nltk.translate.bleu_score import corpus_bleu
3 references = [['GugelhupfLjudski.txt']]
4 candidates = ['GugelhupfGoogle.txt']
5 score = corpus_bleu(references, candidates)
6 print(score)
```

Slika 7 - Prikaz koda za `corpus_bleu()` - recept - Google

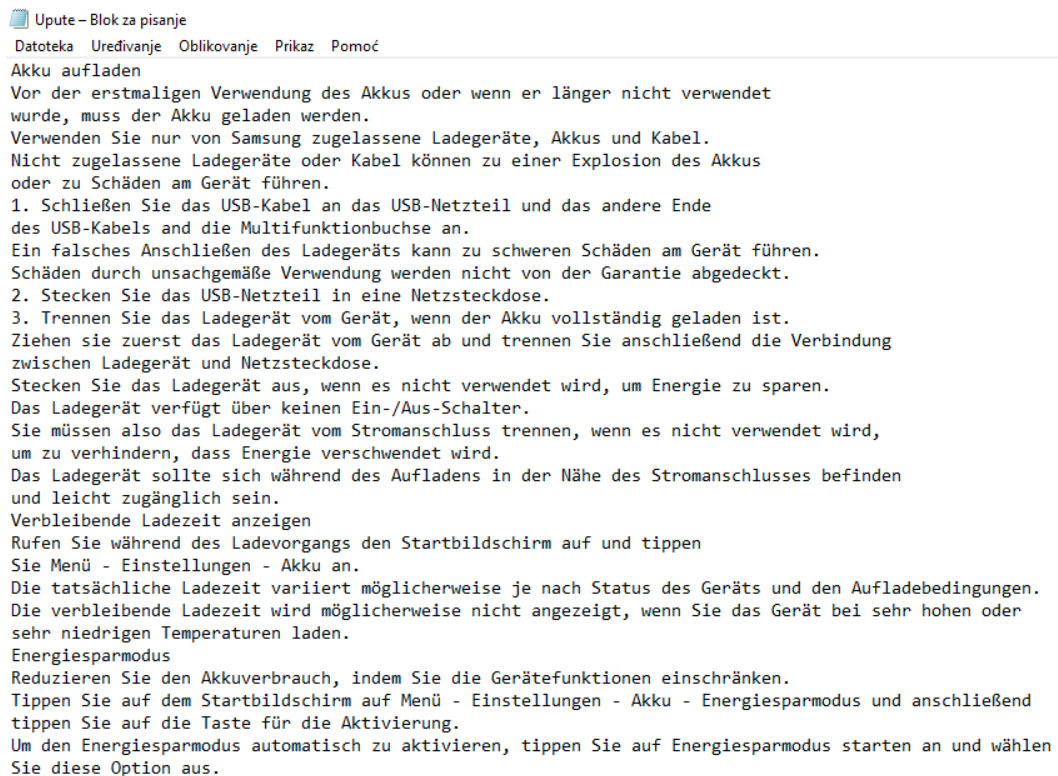
```
In [1]: runfile('C:/Users/Kristina/.spyder-py3/temp.py', wdir='C:/Users/
Kristina/.spyder-py3')
0.5292523437057239
```

Slika 8 - Rezultat `corpus_bleu()` - recept - Google

Iz rezultata je vidljivo kako je Bing online prevoditelj dao bolji rezultat od Google prevoditelja, iako je razlika u konačnim rezultatima mala.

3.2. Upute

Obraden je tekst iz uputa za Samsung mobilne uređaje koji se odnosi na punjenje i štednju baterije.



Slika 9 - Prikaz teksta - upute

3.2.1. Corpus BLEU rezultati – upute

Slijede BLEU rezultati prijevoda dvaju online prevoditelja.

Bing Microsoft Translator

```
2 from nltk.translate.bleu_score import corpus_bleu
3 references = [['UputeLjudski.txt']]
4 candidates = ['UputeBing.txt']
5 score = corpus_bleu(references, candidates)
6 print(score)
7
```

Slika 10 - Prikaz koda za corpus_bleu() - upute – Bing

```
In [1]: runfile('C:/Users/Kristina/.spyder-py3/temp.py', wdir='C:/Users/
Kristina/.spyder-py3')
0.3948627814210998
```

Slika 11 - Rezultat corpus_bleu() - upute - Bing

Google Translate

```
2 from nltk.translate.bleu_score import corpus_bleu
3 references = [['UputeLjudski.txt']]
4 candidates = ['UputeGoogle.txt']
5 score = corpus_bleu(references, candidates)
6 print(score)
7
```

Slika 12 - Prikaz koda za corpus_bleu() - upute - Google

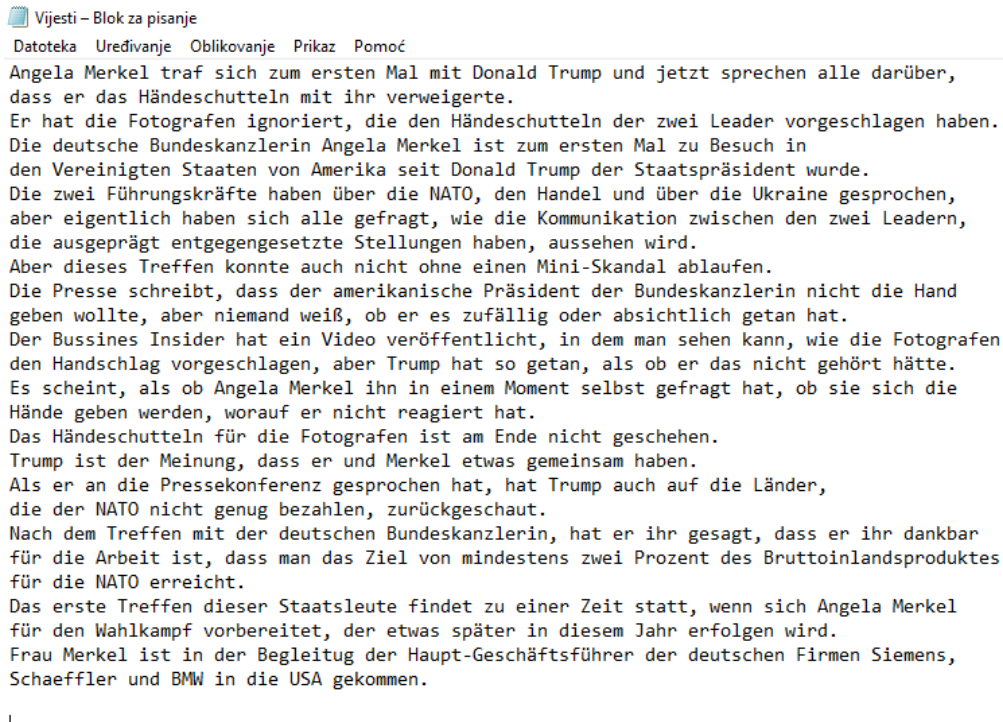
```
In [2]: runfile('C:/Users/Kristina/.spyder-py3/temp.py', wdir='C:/Users/
Kristina/.spyder-py3')
0.38553961859614977
```

Slika 13 - Rezultat corpus_bleu() - upute - Google

Kao i u prethodnom slučaju, tako i ovome Bing prevoditelj daje bolje rezultate od Google prevoditelja.

3.3. Vijesti

Obrađen je medijski članak na temu susreta njemačke kancelarke Angele Merkel i predsjednika SAD-a, Donalda Trumpa.



Slika 14 - Prikaz teksta - Vijesti

3.3.1. Corpus BLEU rezultati – vijesti

Bing Microsoft Translator

```
2 from nltk.translate.bleu_score import corpus_bleu
3 references = [['VijestiLjudski.txt']]
4 candidates = ['VijestiBing.txt']
5 score = corpus_bleu(references, candidates)
6 print(score)
7
```

Slika 15 - Prikaz koda za corpus_bleu() - vijesti - Bing

```
In [1]: runfile('C:/Users/Kristina/.spyder-py3/temp.py', wdir='C:/Users/
Kristina/.spyder-py3')
0.47717543743980667
```

Slika 16 - Rezultat corpus_bleu() - vijesti - Bing

Google Translate

```
2 from nltk.translate.bleu_score import corpus_bleu
3 references = [['VijestiLjudski.txt']]
4 candidates = ['VijestiGoogle.txt']
5 score = corpus_bleu(references, candidates)
6 print(score)
7
```

Slika 17 - Prikaz koda za corpus_bleu() - vijesti - Google

```
In [1]: runfile('C:/Users/Kristina/.spyder-py3/temp.py', wdir='C:/Users/
Kristina/.spyder-py3')
0.46793122417294464
```

Slika 18 - Rezultat corpus_bleu() - vijesti – Google

Bing online prevoditelj i u ovom, kao i u prethodnim slučajevima ima malu prednost nad Google online prevoditeljem.

3.4. BLEU rezultati - zaključak

Iako se čak i letimičnim pregledom dobivenih prijevoda tekstova može uvidjeti kako je Google Translate online prevoditelj dao kvalitetnije prijevode od Bing Microsoft Translator prevoditelja, BLEU metrika daje suprotne rezultate. Prema konačnim rezultatima, Bing prevoditelj je za svaku vrstu teksta, prema BLEU metrici, dao bolje i točnije prijevode (Tablica 1). Već iz toga možemo zaključiti kako BLEU metrika nije posve sigurna i precizna za evaluiranje tekstova prevedenih s jednog jezika na drugi.

BLEU metrika zapravo mjeri izravnu povezanost između riječi referentnog teksta i riječi dobivenih pomoću nekog (strojnog) prevoditelja te u kolikoj su mjeri riječi ili skupine riječi iz navedenih tekstova identične. Samim time, prijevodi koji su u stvarnosti točniji i precizniji mogu postići lošije rezultate dobivene korištenjem BLEU metrike.

BLEU metrika ne može prepoznati fraze ili sinonime, zbog čega također prijevodi koji su u stvarnosti kvalitetniji mogu postići lošije rezultate.

| 1. tekst: Recept | | 2. tekst: Upute | | 3. tekst: Vijesti | |
|------------------|--------|-----------------|--------|-------------------|--------|
| Bing | Google | Bing | Google | Bing | Google |
| 0.5381 | 0.5293 | 0.3949 | 0.3855 | 0.4772 | 0.4679 |

Tablica 1- prikaz rezultata BLEU metrike

4. Character n-gram F-score

ChrF score je metrika za automatsku evaluaciju. Slična je F-mjeri koja se bazira na morfemima, a također u obzir uzima morfo-sintaktičke fenomene. Osim toga, ne zahtijeva nikakve dodatne alate za izračune te je potpuno neovisna o jeziku, a tokenizacija nije potrebna (Popović 2015:1).

Formula za *chrF score* je:

$$CHRF_{\beta} = (1 + \beta^2) \times \frac{CHRP \times CHRR}{\beta^2 \times CHRP + CHRR}$$

gdje je CHRP postotak n-grama u hipotezi (strojnom prijevodu) koji su prisutni u referentnom prijevodu, dok je CHRR postotak n-grama u referentnom prijevodu koji su također prisutni i u hipotezi tj. strojnom prijevodu. β je parametar koji dodjeljuje β puta više važnosti odazivu nego preciznosti, a ukoliko je $\beta = 1$, tada odaziv i preciznost imaju istu važnost.

Prema autorici Popović, dodatni eksperimenti pokazali su da dodavanje 1-grama i 2-grama riječi standardnom *chrF score-u* poboljšava korelaciju s izravnim ocjenjivanjem (procjenama), iako još nije posve jasno koja od tih dviju opcija je bolja – korištenje samo 1-grama (*chrF+*) ili korištenje 1-grama i 2-grama (*chrF++*)

4.1. chrF++

ChrF++ alat je alat za automatsku evaluaciju rezultata strojnog prijevoda, koji se temelji na preciznosti n-grama znakova i njihovom odazivu.

Alat izračunava prosječnu F-mjeru (harmonijsku sredinu preciznosti i odaziva) na svim znakovima i riječima n-grama, gdje je zadani red n-grama znakova 6, a red n-grama riječi 2. Aritmetička sredina se također koristi kako bi se izračunao prosjek n-grama.

ChrF++ alat koristi programski jezik Python te zahtijeva ulaze koji su *reference* i *hypothesis*, odnosno referentni tekst te tekst strojnog prijevoda. Tekstovi moraju biti oblikovani na način da je svaka rečenica smještena u svoj redak – jedna rečenica po liniji.³

Skripta `chrF++.py` korištena je pomoću Naredbenog retka (Command Prompta) na slijedeći način:

```
python chrF++.py -R referentniTekst -H strojniPrijevod
```

Korištenjem navedene skripte dobiju se sljedeći rezultati (output):

1. početno vrijeme
2. ukupna F-mjera na razini dokumenta
3. aritmetička sredina na razini rečenica
4. završno vrijeme

U ovome diplomskom radu korišten je i promijenjeni redoslijed n-grama znakova (sa šest na četiri) te je dan i prikaz rezultata na razini svake pojedine rečenice. F-mjera obuhvaća vrijednosti u rasponu od 0 do 100.

4.2. ChrF++ - rezultati

Slijede prikazi rezultata istih tekstova korištenih prilikom mjerenja rezultata BLEU metrike – recept, upute za Samsung mobilne uređaje te vijesti. Dana je usporedba oba online prevoditelja.

³ Izvor: <https://github.com/m-popovic/chrF/blob/master/README.md>

4.2.1. Rezultati – Tablica

Tablica 2 prikazuje rezultat chrF++ metrike za sve tri vrste teksta.

Retci tablice prikazuju ukupnu F-mjeru na razini dokumenta za red n-grama znakova 6 i 4 (c6+w2-F i c4+w2-F2) te aritmetičku sredinu na razini rečenica, također za red n-grama znakova 6 i 4 (c6+w2-avgF2 i c4+w2-avgF2). Potom su prikazani rezultati ukupne F-mjere za svaku od rečenica pojedine vrste teksta.

| Rezultati | 1. tekst: Recept | | 2. tekst: Upute | | 3. tekst: Vijesti | |
|---------------------|------------------|---------|-----------------|---------|-------------------|---------|
| | Bing | Google | Bing | Google | Bing | Google |
| c6+w2-F2 | 45,1461 | 47,3624 | 54,8492 | 61,9476 | 62,8065 | 66,4684 |
| c6+w2-avgF2 | 43,8202 | 47,0480 | 53,1090 | 58,6607 | 63,4638 | 67,5007 |
| c4+w2-F2 | 50,3173 | 52,0623 | 58,5944 | 64,7410 | 66,6217 | 70,0634 |
| c4+w2-avgF2 | 48,8514 | 51,6090 | 56,7384 | 61,4550 | 67,0887 | 70,7886 |
| 1::c6+w2-F2 | 50,1095 | 67,474 | 100,0000 | 30,5623 | 53,4918 | 58,6231 |
| 2::c6+w2-F2 | 52,6714 | 53,5844 | 62,4937 | 52,1779 | 63,8517 | 65,5697 |
| 3::c6+w2-F2 | 39,9258 | 31,2758 | 64,4933 | 80,0722 | 66,5475 | 64,2743 |
| 4::c6+w2-F2 | 50,177 | 61,7625 | 63,9962 | 68,6104 | 70,4434 | 60,0883 |
| 5::c6+w2-F2 | 35,9446 | 36,9894 | 55,0468 | 68,8478 | 50,6629 | 83,9179 |
| 6::c6+w2-F2 | 49,344 | 49,0164 | 40,1878 | 61,4907 | 77,7645 | 82,7294 |
| 7::c6+w2-F2 | 29,0763 | 32,0202 | 37,2871 | 67,6736 | 56,6806 | 50,4534 |
| 8::c6+w2-F2 | 48,5883 | 62,4525 | 43,8519 | 58,6595 | 54,4086 | 71,8779 |
| 9::c6+w2-F2 | 26,1776 | 25,3327 | 64,5895 | 64,6221 | 53,0433 | 46,3702 |
| 10::c6+w2-F2 | 47,684 | 51,1364 | 53,4415 | 51,5391 | 100,0000 | 82,5395 |
| 11::c6+w2-F2 | 40,1437 | 43,7148 | 25,085 | 74,8622 | 74,7717 | 78,7379 |
| 12::c6+w2-F2 | 11,6149 | 35,7613 | 31,8757 | 34,5453 | 40,2939 | 52,9864 |
| 13::c6+w2-F2 | 65,3154 | 64,7448 | 58,3273 | 78,2641 | 64,6816 | 66,1576 |
| 14::c6+w2-F2 | 25,6858 | 23,2056 | 46,8008 | 63,0411 | 61,8511 | 80,6846 |
| 15::c6+w2-F2 | 27,0699 | 27,1713 | 45,8329 | 48,1447 | - | - |
| 16::c6+w2-F2 | 55,7961 | 56,1689 | 58,5807 | 76,3179 | - | - |
| 17::c6+w2-F2 | 66,8136 | 54,5248 | 89,0377 | 89,0377 | - | - |
| 18::c6+w2-F2 | 66,6252 | 70,5275 | 71,96 | 71,9892 | - | - |
| 19::c6+w2-F2 | - | - | 4,4681 | 13,4187 | - | - |
| 20::c6+w2-F2 | - | - | 60,3563 | 60,3563 | - | - |

| | | | | | | |
|---------------------|---|---|---------|---------|---|---|
| 21::c6+w2-F2 | - | - | 49,8702 | 36,8777 | - | - |
| 22::c6+w2-F2 | - | - | 40,8155 | 39,4253 | - | - |

Tablica 2- prikaz rezultata chrF++ metrike

Za prvu vrstu teksta, odnosno tekst „Recept“, prema chrF++ metrici Google prevoditelj dao je bolje rezultate od Bing prevoditelja, što je u skladu s rezultatima ljudske evaluacije.

Također i u drugom slučaju, za vrstu teksta „Upute“, rezultati chrF++ metrike prikazuju kako je Google prevoditelj dao uspješniji tj. bolji prijevod od Bing prevoditelja.

Za vrstu teksta „Vijesti“, kao i prethodnim slučajevima, rezultati idu u korist Google prevoditelja, što se za razliku od rezultata BLEU metrike podudara s rezultatima ljudske evaluacije.

5. Ljudska evaluacija na temelju MQM-a

Ljudska evaluacija, za razliku od automatske evaluacije koja koristi formule i izračune, je evaluacija kojom se pomoću ocjenjivača (osobe) vrednuje prijevod.

Kao glavni kriteriji pri ocjenjivanju prijevoda koriste se tečnost i adekvatnost.

Tečnost se odnosi na ispravnost prijevoda s gramatičke razine, dok se adekvatnost odnosi na značenje samoga prijevoda, primjerice na dio teksta u koji je izostavljen u prijevodu (*omission*), na pogrešan prijevod (*mistranslation*), na dio teksta koji je ostao nepreveden tj. jednak obliku riječi na jeziku koji se prevodi (*untranslated*) te na dio teksta koji je dodan, odnosno koji se ne nalazi u tekstu jezika koji se prevodi (*addition*) i sl.

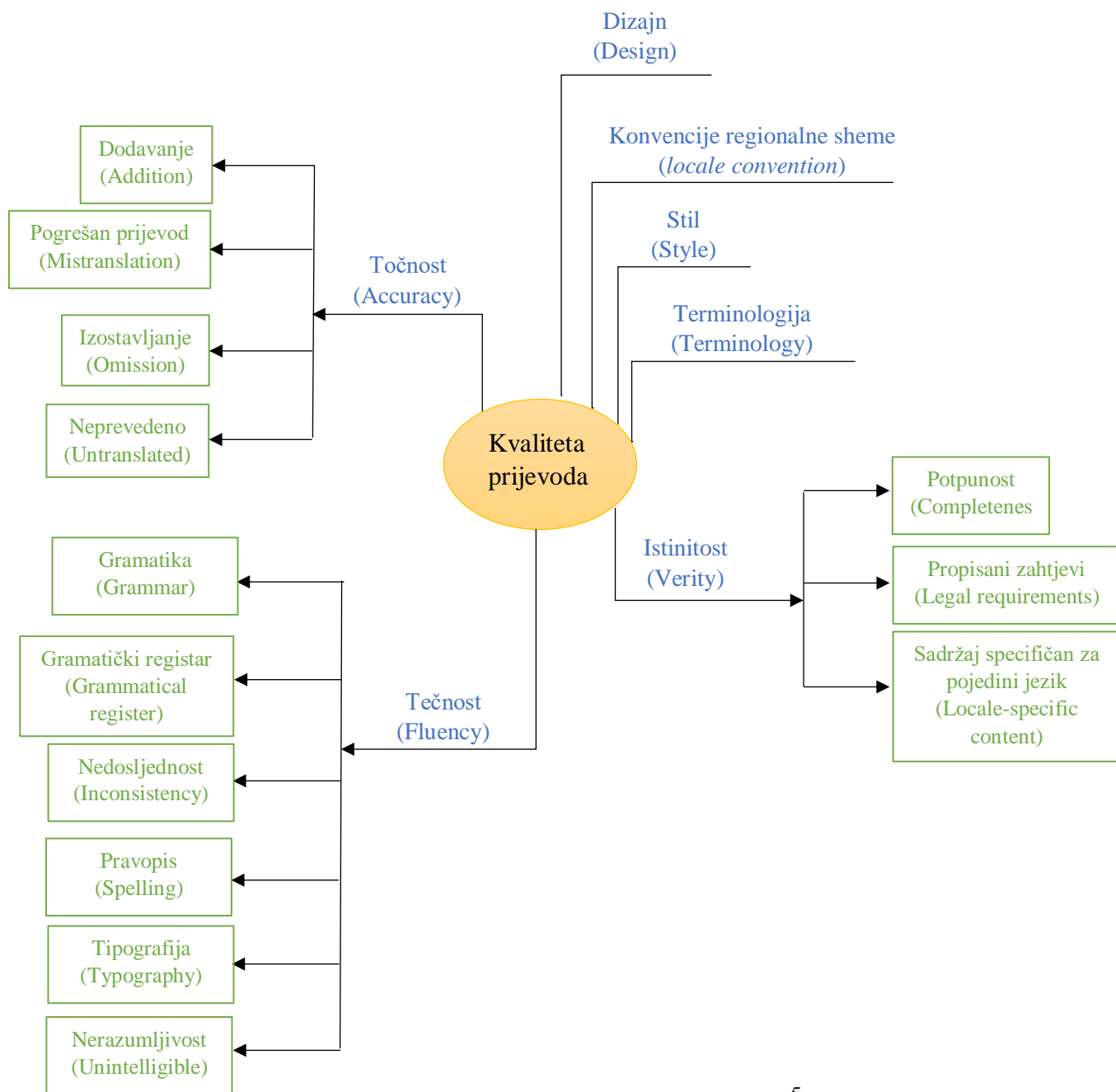
5.1. MQM metrika

MQM ili *Multidimensional Quality Metrics* je metrika koja se koristi za ocjenu kvalitete prijevoda.⁴ Njome se definiraju standardna značenja kategorija grešaka (engl. *issue types*).

Ova višedimenzionalna metrika obuhvaća hijerarhijski popis kategorija pogrešaka prijevoda. Glavne kategorije se nadalje dijele na dodatne podkategorije kako bi se čim specifičnije mogla odrediti vrsta pogreške.

Glavne kategorije su točnost (*accuracy*), tečnost (*fluency*), dizajn (*design*), konvencije regionalne sheme (*locale convention*), stil (*style*), terminologija (*terminology*) te istinitost (*verity*).

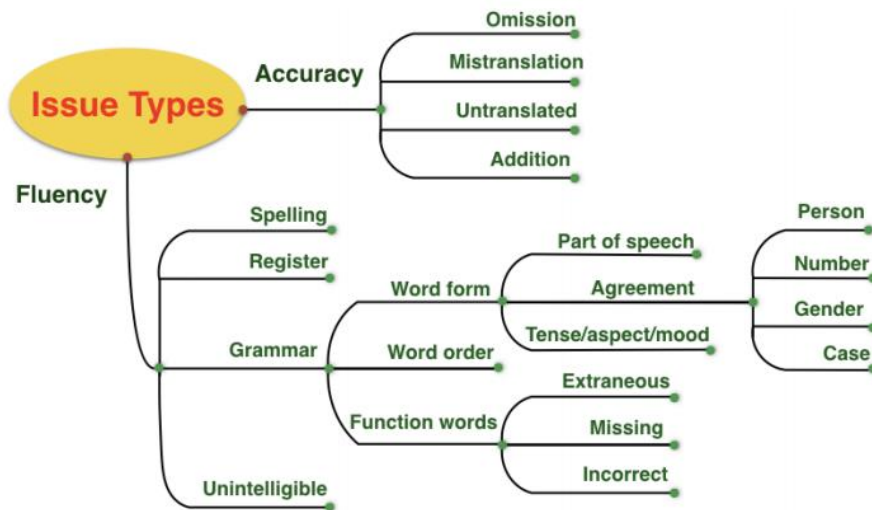
⁴ Izvor: <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>



Slika 19 - Prikaz MQM jezgre⁵

⁵ Slika prilagođena korištenjem izvora: <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>

Prema primjeru autora Klubička, Toral i Sánchez-Cartagena, za izradu ovoga seminarskog rada korištena je MQM metrika sa skupom oznaka za slavenske jezike, uz dodatnu preinaku – umjesto korištenja kategorije „register“ zadržana je kategorija „typography“:



Slika 20 - MQM - Slavic tagset

5.2. Rezultati ljudske evaluacije

Za evaluaciju prijevoda pomoću MQM metrike, korišten je TREAT alat koji omogućava označavanje pogrešaka prema kategorijama, a kao rezultat daje statistiku pogrešaka te vizualnu statistiku u obliku linijskog i tortnog grafikona.

Evaluacija je provedena nad istim tekstovima kao i BLEU metrika i chrF++.

Rezultati su prikazani pomoću linijskog i tortnog grafikona za svaki od online prevoditelja.

Prilikom ljudske evaluacije u alatu TREAT korišten je i referentni prijevod, ali je važno napomenuti kako se nisu označavale pogreške s obzirom na referencu, odnosno nije se tražio potpuno identičan prijevod, već je referentni prijevod služio samo kao dodatna pomoć pri ocjenjivanju strojnog prijevoda.

5.2.1. Ljudska evaluacija – tablica

U tablici 3 su prikazani konačni rezultati ljudske evaluacije – ukupan broj rečenica, ukupan broj pogrešaka, prosječan broj pogrešaka po rečenici te broj pogrešaka za svaku od kategorija.

| Rezultati | 1. tekst: Recept | | 2. tekst: Upute | | 3. tekst: Vijesti | |
|--------------------------------------|------------------|--------|-----------------|--------|-------------------|--------|
| | Bing | Google | Bing | Google | Bing | Google |
| Ukupan broj rečenica | 18 | 18 | 22 | 22 | 14 | 14 |
| Ukupan broj pogrešaka | 66 | 58 | 35 | 24 | 28 | 22 |
| Prosječan broj pogrešaka po rečenici | 3,66 | 3,22 | 1,59 | 1,09 | 2 | 1,57 |
| <i>Accuracy</i> | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Addition</i> | 2 | 1 | 0 | 0 | 0 | 0 |
| <i>Mistranslation</i> | 38 | 29 | 20 | 21 | 20 | 15 |
| <i>Omission</i> | 1 | 2 | 2 | 1 | 2 | 0 |
| <i>Untranslated</i> | 8 | 18 | 0 | 0 | 0 | 0 |
| <i>Fluency</i> | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Grammar</i> | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Word form</i> | 5 | 2 | 0 | 0 | 0 | 0 |
| <i>Part of speech</i> | 1 | 0 | 0 | 0 | 0 | 0 |
| <i>Agreement</i> | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Person</i> | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Number</i> | 0 | 0 | 0 | 0 | 2 | 0 |
| <i>Gender</i> | 0 | 1 | 0 | 0 | 1 | 2 |
| <i>Case</i> | 2 | 2 | 2 | 0 | 0 | 1 |
| <i>Tense/aspect/mood</i> | 0 | 0 | 0 | 0 | 2 | 4 |
| <i>Word order</i> | 6 | 1 | 3 | 0 | 1 | 0 |
| <i>Function words</i> | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Extraneous</i> | 1 | 0 | 0 | 0 | 0 | 0 |
| <i>Missing</i> | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Incorrect</i> | 1 | 1 | 3 | 2 | 0 | 0 |

| | | | | | | |
|-----------------------|---|---|---|---|---|---|
| <i>Spelling</i> | 0 | 1 | 5 | 0 | 0 | 0 |
| <i>Typography</i> | 0 | 0 | 0 | 0 | 0 | 0 |
| <i>Unintelligible</i> | 1 | 0 | 0 | 0 | 0 | 0 |

Tablica 3 - Prikaz rezultata ljudske evaluacije prema MQM metrici

Iz statističkih rezultata vidljivo je kako je Google prevoditelj za tekst „Recept“ dao manje pogrešaka (58) od Bing prevoditelja (66).

Za tekst „Upute“ Google prevoditelj je ponovno dao bolje rezultate tj. bolji prijevod. Google prevoditelj dao je 24 pogreške, dok je Bing prevoditelj dao 35 pogrešaka.

Prijevod teksta iz područja vijesti također je uspješnije izvršen pomoću Google prevoditelja koji je dao 22 pogreške, dok je Bing prevoditelj dao 28 pogrešaka.

U sva tri slučaja (sve tri vrste teksta) najčešća kategorija pogrešaka je pogrešan prijevod, odnosno *mistranslation* koja spada u glavnu kategoriju točnosti, dok se ostale pogreške uglavnom odnose na one gramatičke vrste, primjerice podudaranje u padežu i rodu te dodavanje funkcijskih riječi (zamjenica, priloga, prijedloga...) ili pak njihovo izostavljanje (*missing*) ili pogrešnu uporabu tj. prijevod (*incorrect*).

5.3. Usporedba pogrešaka strojnih prijevoda

Ovo potpoglavlje prikazuje neke od konkretnih pogrešaka u alatu TREAT. Dan je opis i usporedba dvaju online prevoditelja.

5.3.1. Vrsta teksta – recept

Kao što je već navedeno, najčešće pogreške spadaju u kategoriju pogrešnog prijevoda, odnosno *mistranslation* pa je tako primjerice „dizano tijesto“ prevedeno kao „klica tijesto“ (slika 21).

Source sentence:

Er kann sowohl aus Germteig, Rührteig oder Biskuiteig hergestellt werden.

Reference sentence:

Može se napraviti od dizanog tijesta, lijevanog tijesta ili biskvita.

System sentence:

Current annotation:Mistranslation

Može se napraviti i od **klice** tijesta, **miješanja** tijesta ili **spužve** tijesta.

Slika 21 - Treat - recept – Bing 1

S druge pak strane, u ovoj vrsti teksta, Bing prevoditelj niti jedan put njemačku riječ „Gugelhupf“ nije preveo kao „kuglof“, već je ona ostala u svom izvornom obliku (slika 22).

Source sentence:

Auch in pikanter Form ist der Gugelhupf ein Genuss.

Reference sentence:

Čak i u slanom obliku je kuglof užitak.

System sentence:

Current annotation:Untranslated

Čak i u začinjenom obliku, **Gugelhupf** je poslastica.

Slika 22 - Treat - recept - Bing 2

Slika 23 prikazuje rečenicu označenu kao *unintelligible* odnosno rečenica je nerazumljiva. Iako bi se svaka pogreška mogla označiti svojom pripadajućom kategorijom, prema MQM

smjernicama⁶, ova rečenica sadrži dovoljan broj pogrešaka da bi se mogla pripisati kategoriji *unintelligible*.

Source sentence:

Die geschmacklichen Varianten reichen von Marmorgugelhupf, Gugelhupf mit Rosinen, Kürbiskerngugelhupf bis zu Vollkorngugelhupf oder beschwipsten Versionen wie Glühweingugelhupf und Eierlikörgugelhupf.

Reference sentence:

Različiti okusi kreću se od mramornog kuglofa, kuglofa s grožđicama, kuglofa sa sjemenkama bundeve pa sve do kuglofa s integralnim brašnom ili pijanih verzija kao što su kuglof sa kuhanim vinom ili sa likerom od jaja.

System sentence:

Current annotation:Unintelligible

Ukusna varijante u rasponu od mramora kuglof, kuglof s grožđicama, bundeve sjeme kuglof na čežitarice kuglof ili hrane verzije kao što su kuhano vino kuglof i punč kuglof.

Slika 23 - Treat - recept - Bing 3

Četvrti pak primjer, odnosno slika 24, ponovno prikazuje kategoriju pogreške *mistranslation*, gdje je riječ „Form“ prevedena kao plijesan, iako je njemački termin za plijesan „der Schimmelpilz“. Međutim, prema nekim izvorima (prema nekolicini ostalih online prevoditelja) se kao njemački termin za plijesan pojavljuje i termin „Form“. Postoji mogućnost da je ta riječ zastarjelica te da se više ne koristi budući da se na stranici Duden⁷ koja obuhvaća riječnik njemačkog jezika, a ujedno i objašnjenje vokabulara, kao objašnjenje pojma „die Form“ ne nalazi pojam koji objašnjava pojam plijesni. U suprotnom dolazimo do problema nerazumijevanja konteksta koji se javlja kod strojnih prevoditelja.

Source sentence:

Dann die Form vorsichtig abziehen.

Reference sentence:

Tada oprezno skinite kalup.

System sentence:

Current annotation:Mistranslation

Zatim nježno **oguliti plijesan.**

Slika 24 - Treat - recept - Bing 4

⁶ Izvor: <http://www.qt21.eu/downloads/annotatorsGuidelines-2014-06-11.pdf>

⁷ Link: <https://www.duden.de/>

Na slikama 25 i 26 prikazano je kako ni Google prevoditelj njemačku riječ „Gugelhupf“ nije preveo kao „kuglof“, već ju je također ostavio u njenom izvornom obliku, što se događa kroz cijeli tekst. U prvom primjeru (slika 25) je dodatna pogreška prijevod „torta“, iako riječ „der Kuchen“ pored „kolača“ znači i „torta“, međutim ne možemo reći da je kuglof vrsta torte, stoga je pogreška označena kao *mistranslation* te opet dolazimo do problema nerazumijevanja konteksta.

Source sentence:

Der Kuchen mit der wohl charakteristischsten Form ist der Gugelhupf.

Reference sentence:

Kolač sa vjerojatno najkarakterističnijim oblikom je kuglof.

System sentence:

Current annotation:Untranslated

Torta s vjerojatno najkarakterističnijim oblikom je **Gugelhupf**.

Slika 25 - Treat - recept - Google 1

Source sentence:

Die geschmacklichen Varianten reichen von Marmorgugelhupf, Gugelhupf mit Rosinen, Kürbiskerngugelhupf bis zu Vollkorngugelhupf oder beschwipsten Versionen wie Glühweingugelhupf und Eierlikörgugelhupf.

Reference sentence:

Različiti okusi kreću se od mramornog kuglofa, kuglofa s grožđicama, kuglofa sa sjemenkama bundeve pa sve do kuglofa s integralnim brašnom ili pijanih verzija kao što su kuglof sa kuhanim vinom ili sa likerom od jaja.

System sentence:

Current annotation:Untranslated

Varijacije okusa kreću se od **kuglice** od mramorne **kugle, Gugelhupfa** s grožđicama, **Kürbiskerngugelhupf** do **Vollkorngugelhupf** ili pijanih verzija kao što su **Glühweingugelhupf** i **Eierlikörgugelhupf**.

Slika 26 - Treat - recept - Google 2

5.3.2. Vrsta teksta – upute

Slika 27 prikazuje primjer pogreške *omission*, odnosno u strojnom prijevodu nedostaje ključna riječ koja se nalazi unutar teksta na stranome jeziku. Izostavljanje riječi prikazano je označavanjem razmaka između dviju riječi.

Source sentence:

Das Ladegerät verfügt über keinen Ein-/Aus-Schalter.

Reference sentence:

Punjač ne raspolaže prekidačem za paljenje / gašenje.

System sentence:

Current annotation: Omission

Punjač nema prekidač za isključivanje.

Slika 27 - Treat - upute - Bing 1

Slika 28 prikazuje pogrešan prijevod, gdje je „Energiesparmodus“ koji se osim „Štednja baterije“ može prevesti i kao „Način rada za uštedu energije“, preveden kao glagol „spavati“.

Source sentence:

Energiesparmodus

Reference sentence:

Štednja baterije

System sentence:

Current annotation: Mistranslation

Spavati

Slika 28 - Treat - upute - Bing 2

Može se reći da je Google prevoditelj podosta uspješno preveo tekst „Upute“. Veliki broj rečenica nije označen niti jednom pogreškom, što je pomalo iznenađujuće.

Greške koje se javljaju su najčešće one kategorije *mistranslation*, kao što je to u ova dva primjera (slika 29 i slika 30). Na slici 30 vidimo kako je Google prevoditelj dijelove teksta na njemačkom jeziku preveo na engleski jezik, umjesto na hrvatski.

Source sentence:

Verbleibende Ladezeit anzeigen

Reference sentence:

Preostalo vrijeme punjenja

System sentence:

Current annotation: Mistranslation

Prikaži preostalo vrijeme **učitavanja**

Slika 29 - Treat - upute - Google 1

Source sentence:

Tippen Sie auf dem Startbildschirm auf Menü - Einstellungen - Akku - Energiesparmodus und anschließend tippen Sie auf die Taste für die Aktivierung.

Reference sentence:

Na početnom zaslon dodirnite Meni - Postavke - Baterija - Štednja baterije i zatim dodirnite prekidač za aktivaciju.

System sentence:

Current annotation: Mistranslation

Na početnom zaslonu **taknite Menu - Settings - Battery - Sleep mode**, a zatim dodirnite **gumb Activation**.

Slika 30 - Treat - upute - Google 2

5.3.3. Vrsta teksta – vijesti

Prvi primjer prijevoda Bing prevoditelja (slika 31) prikazuje nekoliko pogrešaka različitih kategorija. Tako primjerice osim kategorije pogrešnog prijevoda imamo i pogrešku glagolskog vremena (prijevod glasi „ne može“, umjesto perfekta „nije mogao“) te pogrešku poretka tj. redoslijeda riječi (*word order*), gdje je riječ „također“ smještena na kraj rečenice, što nije u duhu hrvatskoga jezika.

Source sentence:

Aber dieses Treffen konnte auch nicht ohne einen Mini-Skandal ablaufen.

Reference sentence:

No ni ovaj sastanak nije mogao proći bez mini skandala.

System sentence:

Current annotation:Word order

Ali **taj** sastanak **ne može isteći** bez mini skandala, **također**.

Slika 31 - Treat - vijesti - Bing 1

Slika 32 prikazuje problem korištenja strojnih prevoditelja kada recimo imamo naziv neke tvrtke, novina i sl. gdje se njihov naziv ne prevodi, dakle točan prijevod u ovome slučaju bio bi, kao što je prikazano u referentnom prijevodu, „Bussines Insider“. Druga pogreška je prijevod „možete“ koja se odnosi na dio rečenice „in dem man sehen kann“ – na njemačkom jeziku „man“ se prevodi kao „se“ pa bi stoga prijevod trebao biti „se može vidjeti“, umjesto „možete vidjeti“. Naravno, hoće li se to uzimati kao pogreška ovisi o ljudskom evaluatoru.

Source sentence:

Der Bussines Insider hat ein Video veröffentlicht, in dem man sehen kann, wie die Fotografen den Handschlag vorgeschlagen, aber Trump hat so getan, als ob er das nicht gehört hätte.

Reference sentence:

Bussines Insider je objavio snimku na kojoj se vidi kako fotografi sugeriraju rukovanje, ali Trump se pravio kao da je to nije čuo.

System sentence:

Current annotation:Mistranslation

Insajder bussines je objavio video u kojem **možete** vidjeti fotografe koji sugeriraju rukovanje, ali Trump se pretvarao da to nije čuo.

Slika 32 - Treat - vijesti - Bing 1

Slika 33 prikazuje slučaj jednak onome pri korištenju Bing prevoditelja – naziv se ne prevodi te treba ostati „Bussines Insider“, dok slika 34 prikazuje pogreške u prijevodu koje pripadaju gramatičkoj kategoriji – odnosno rodu i padežu.

Source sentence:

Der Bussines Insider hat ein Video veröffentlicht, in dem man sehen kann, wie die Fotografen den Handschlag vorgeschlagen, aber Trump hat so getan, als ob er das nicht gehört hätte.

Reference sentence:

Bussines Insider je objavio snimku na kojoj se vidi kako fotografi sugeriraju rukovanje, ali Trump se pravio kao da je to nije čuo.

System sentence:

Current annotation: Mistranslation

Bussineov insajder objavio je videozapis koji pokazuje kako su fotografi predložili rukovanje, ali Trump se pretvarao da to nije čuo.

Slika 33 - Treat - vijesti - Google 1

Source sentence:

Nach dem Treffen mit der deutschen Bundeskanzlerin, hat er ihr gesagt, dass er ihr dankbar für die Arbeit ist, dass man das Ziel von mindestens zwei Prozent des Bruttoinlandsproduktes für die NATO erreicht.

Reference sentence:

Nakon sastanka s njemačkom kancelarkom, kazao je kako joj je zahvalan na radu na tome da se dostigne cilj od barem dva posto BDP-a za NATO.

System sentence:

Current annotation: Gender

Nakon sastanka s njemačkim **kancelarom** rekao joj je da je zahvalan na poslu **koji je postigao** cilj od najmanje dva posto bruto domaćeg proizvoda za NATO.

Slika 34 - Treat - vijesti - Google 2

6. Pearsonov koeficijent korelacije

Korelacije se uglavnom koriste u području statistike i statističkih mjerenja te se njima prikazuju odnosi između dviju varijabli.

U ovome diplomskom radu korištena je Pearsonova korelacija, odnosno Pearsonov koeficijent korelacije koji se koristi u radu s linearnim modelima. Dakle, Pearsonov koeficijent korelacije određuje način na koji veza između pojedinih varijabli može biti opisana linijom.

Formula za izračun korelacije koristi dvije varijable – X i Y, a Pearsonova korelacija označava se sa r :

$$r_{X,Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Ukratko, možemo reći da je Pearsonova korelacija omjer *kovarijanca*⁸ i produkta *standardnih devijacija*⁹.

Formula uvijek daje rezultat u rasponu od -1 do 1:

1. *Pozitivna korelacija* – ukoliko se jedna varijabla povećava, također se povećava i druga varijabla.
2. *Negativna korelacija* – ukoliko se jedna varijabla povećava, druga varijabla se smanjuje.
3. *Nema korelacije* – ukoliko se jedna varijabla povećava, druga varijabla se ne povećava niti se smanjuje – varijable nisu povezane.

Za potrebe diplomskog rada izvršen je višestruki izračun korelacija i to za BLEU i chrF++ metriku u odnosu na ljudsku evaluaciju.

Za svaku vrstu teksta te za svaki od prevoditelja izračunati su BLEU i chrF++ rezultati na razini rečenice. Potom se za svaku rečenicu u korpusu izračunao ukupan broj pogrešaka. Dakle, za izračun Pearsonovog koeficijenta korelacije koristile su se dvije

⁸ Kovarijanca: veličina koja govori o zavisnosti X i Y varijabli.

⁹ Standardna devijacija: označava mjeru raspršenosti u skupu.

varijable za obje metrike: rezultat automatske metrike na razini rečenice te ukupan broj pogrešaka unutar jedne rečenice.

Potom je na temelju rezultata ljudske evaluacije, odnosno pomoću MQM metrike, izračunat Pearsonov koeficijent korelacije i to na način da je za BLEU i chrF++ metriku izračunata korelacija s kategorijom MQM metrike kojoj u korpusu pripada najviše pogrešaka – u ovom slučaju je to kategorija *Accuracy*, a također i za podkategoriju kojoj pripada najviše pogrešaka – *Mistranslation*.

6.1. Pearsonov koeficijent korelacije: BLEU metrika

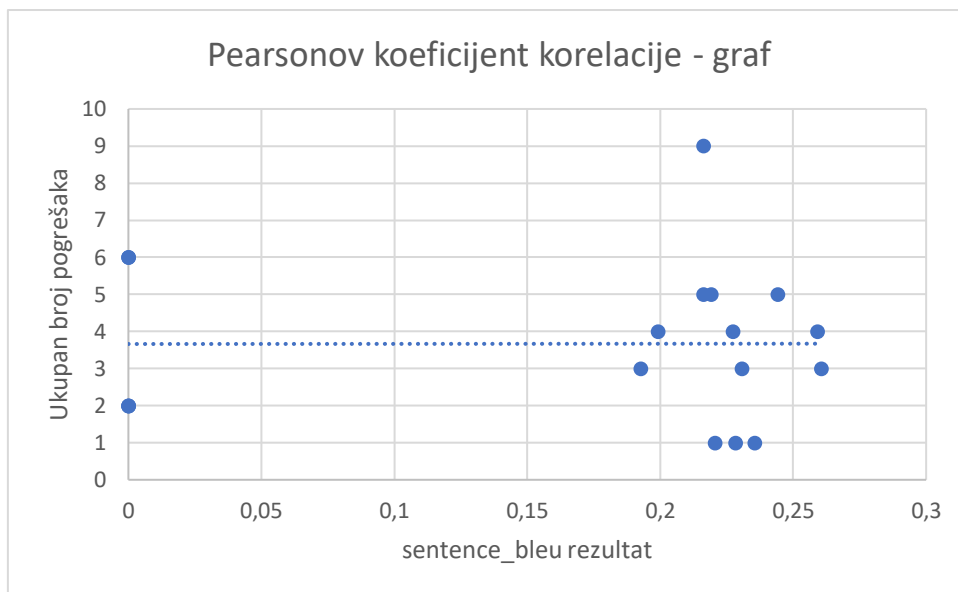
Slijedi prikaz rezultata Pearsonovog koeficijenta dviju varijabli – izračuna BLEU metrike na razini rečenice te ukupnog broja pogrešaka, također na razini rečenice za online prevoditelje – Bing i Google.

6.1.1. Rezultati: Vrsta teksta – Recept

Bing Microsoft Translator

| | sentence_bleu | Ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------------|---------------------|-----------------------|
| 1. | 0,199137491 | 4 | 0,001611393 |
| 2. | 0,19259074 | 3 | |
| 3. | 0,235672694 | 1 | |
| 4. | 0 | 2 | |
| 5. | 0,228368173 | 1 | |
| 6. | 0 | 2 | |
| 7. | 0,216254984 | 9 | |
| 8. | 0 | 6 | |
| 9. | 0 | 6 | |
| 10. | 0,219162941 | 5 | |
| 11. | 0,227285169 | 4 | |
| 12. | 0 | 2 | |
| 13. | 0,230786988 | 3 | |
| 14. | 0,259167113 | 4 | |
| 15. | 0,244234644 | 5 | |
| 16. | 0,216254984 | 5 | |
| 17. | 0,260630129 | 3 | |
| 18. | 0,220690914 | 1 | |

Slika 35 - Pearson - Recept - Bing



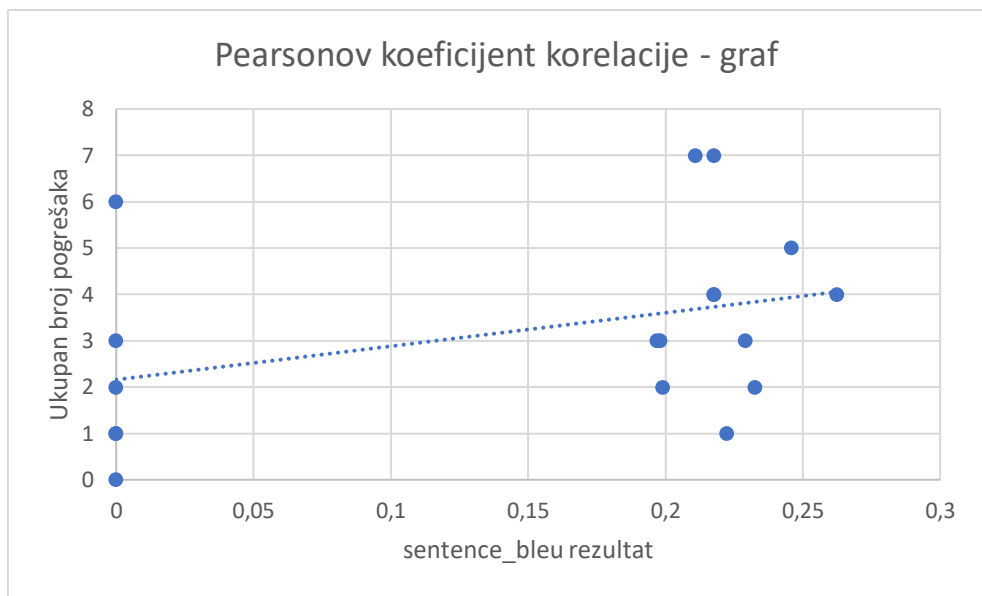
Slika 36 - Pearson - Graf - Recept - Bing

Vidljivo je kako je kako je korelacija skoro pa nepostojeća – rezultat je 0,001611393. Dakle, korelacije između ukupnog broja pogrešaka na razini rečenice i rezultata dobivenog funkcijom *sentence_bleu()* gotovo da i nema.

Google Translate

| | sentence_bleu | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------------|---------------------|-----------------------|
| 1. | 0,199137491 | 2 | 0,382216419 |
| 2. | 0 | 2 | |
| 3. | 0 | 3 | |
| 4. | 0 | 1 | |
| 5. | 0,210988247 | 7 | |
| 6. | 0,196924421 | 3 | |
| 7. | 0,217684669 | 7 | |
| 8. | 0 | 1 | |
| 9. | 0 | 6 | |
| 10. | 0,217684669 | 4 | |
| 11. | 0,229030278 | 3 | |
| 12. | 0 | 0 | |
| 13. | 0,232523241 | 2 | |
| 14. | 0,198176324 | 3 | |
| 15. | 0,245981058 | 5 | |
| 16. | 0,217684669 | 4 | |
| 17. | 0,262447205 | 4 | |
| 18. | 0,222268977 | 1 | |

Slika 37 - Pearson - Recept - Google



Slika 38 - Pearson - Graf - Recept - Google

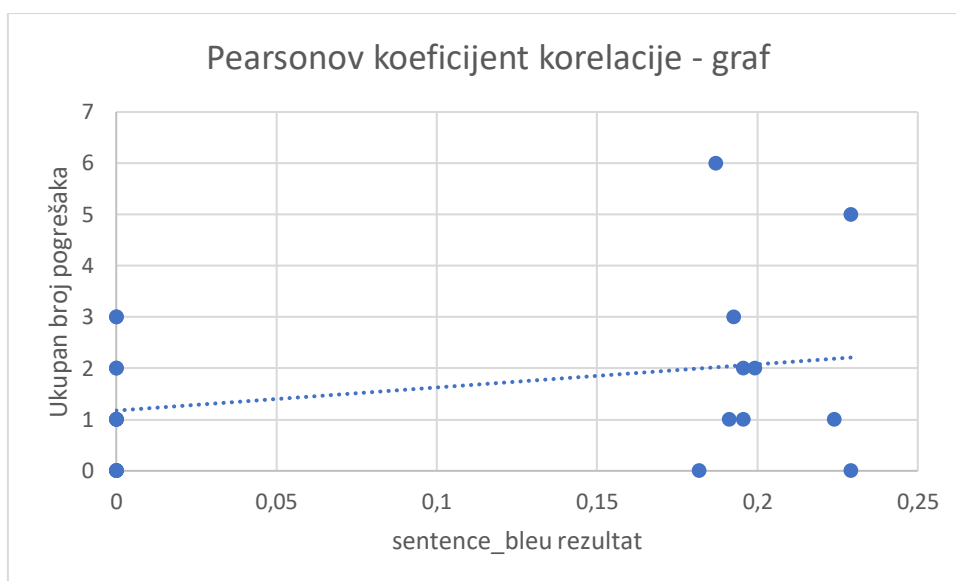
Za razliku od Bing prevoditelja, Google Translate dao je prijevod koji je rezultirao značajnijim koeficijentom korelacije za istu vrstu teksta, 0,382216419 što je pozitivna korelacija – povećanjem vrijednosti jedne varijable, povećava se i vrijednost druge varijable.

6.1.2. Rezultati: Vrsta teksta – Upute

Bing Microsoft Translator

| | sentence_bleu | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------------|---------------------|-----------------------|
| 1. | 0 | 0 | 0,346160958 |
| 2. | 0 | 1 | |
| 3. | 0,19552796 | 1 | |
| 4. | 0 | 3 | |
| 5. | 0,223896339 | 1 | |
| 6. | 0 | 1 | |
| 7. | 0 | 0 | |
| 8. | 0,199137491 | 2 | |
| 9. | 0,19552796 | 2 | |
| 10. | 0,229030278 | 5 | |
| 11. | 0 | 2 | |
| 12. | 0 | 1 | |
| 13. | 0 | 3 | |
| 14. | 0,191123284 | 1 | |
| 15. | 0 | 0 | |
| 16. | 0,19259074 | 3 | |
| 17. | 0,229030278 | 0 | |
| 18. | 0 | 2 | |
| 19. | 0 | 1 | |
| 20. | 0 | 0 | |
| 21. | 0,186905186 | 6 | |
| 22. | 0,181848041 | 0 | |

Slika 39 - Pearson – Upute – Bing



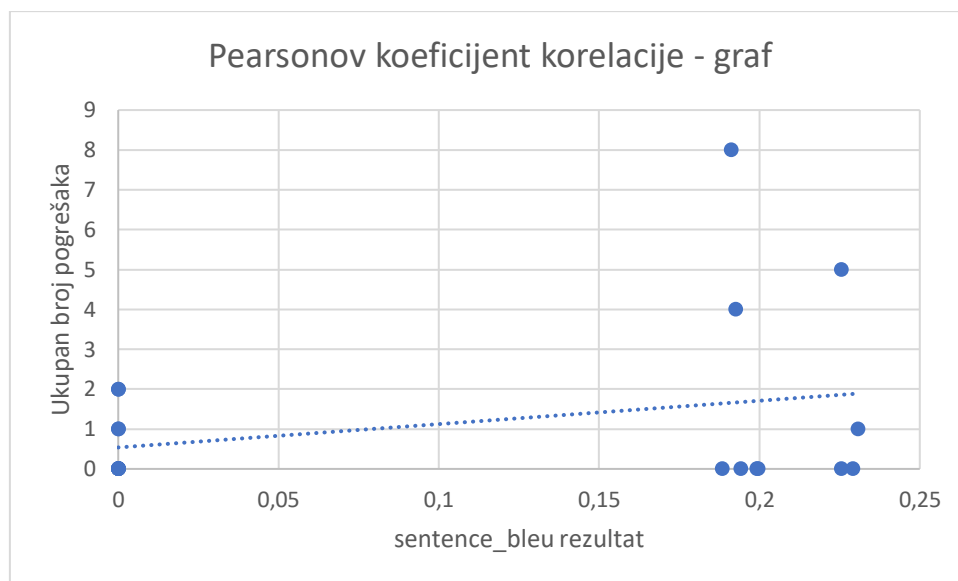
Slika 40 - Pearson – Graf – Upute – Bing

Pearsonov koeficijent korelacije za prijevod teksta „Upute“ dobiven pomoću Bing prevoditelja iznosi 0,346160958, što je pozitivna korelacija.

Google Translate

| | sentence_bleu | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------------|---------------------|-----------------------|
| 1. | 0 | 1 | 0,212577693 |
| 2. | 0 | 2 | |
| 3. | 0,194067951 | 0 | |
| 4. | 0 | 0 | |
| 5. | 0,188273629 | 0 | |
| 6. | 0 | 0 | |
| 7. | 0 | 0 | |
| 8. | 0,199530877 | 0 | |
| 9. | 0 | 2 | |
| 10. | 0,19259074 | 4 | |
| 11. | 0 | 0 | |
| 12. | 0,199137491 | 0 | |
| 13. | 0 | 0 | |
| 14. | 0,225570312 | 0 | |
| 15. | 0 | 1 | |
| 16. | 0,230786988 | 1 | |
| 17. | 0,229030278 | 0 | |
| 18. | 0 | 0 | |
| 19. | 0 | 0 | |
| 20. | 0 | 0 | |
| 21. | 0,191123284 | 8 | |
| 22. | 0,225570312 | 5 | |

Slika 41 - Pearson – Upute – Google



Slika 42 - Pearson – Graf - Upute – Google

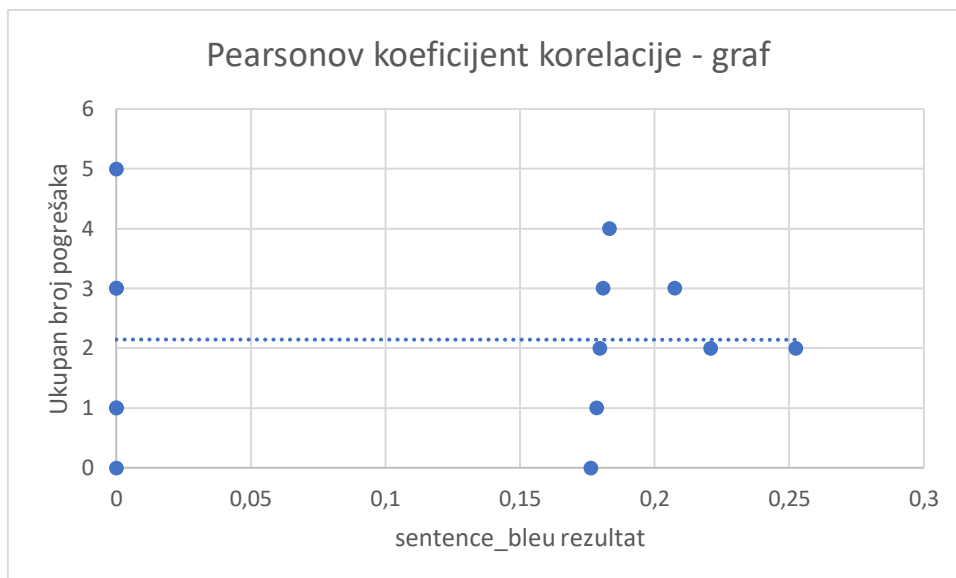
Prijevod dobiven pomoću Google prevoditelja, dao je rezultat koeficijenta korelacije koji iznosi 0,212577693, što znači da je kod Bing prevoditelja veća pozitivna korelacija – kao što je već objašnjeno, povećanjem jedne varijable, povećava se i vrijednost druge.

6.1.3. Rezultati: Vrsta teksta – Vijesti

Bing Microsoft Translator

| | sentence_bleu | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------------|---------------------|-----------------------|
| 1. | 0,252398381 | 2 | -0,001103102 |
| 2. | 0 | 1 | |
| 3. | 0 | 1 | |
| 4. | 0,207459692 | 3 | |
| 5. | 0 | 5 | |
| 6. | 0,179560769 | 2 | |
| 7. | 0,180685379 | 3 | |
| 8. | 0,183050258 | 4 | |
| 9. | 0 | 3 | |
| 10. | 0,176215715 | 0 | |
| 11. | 0 | 0 | |
| 12. | 0 | 3 | |
| 13. | 0,178472607 | 1 | |
| 14. | 0,220690914 | 2 | |

Slika 43 - Pearson - Vijesti – Bing



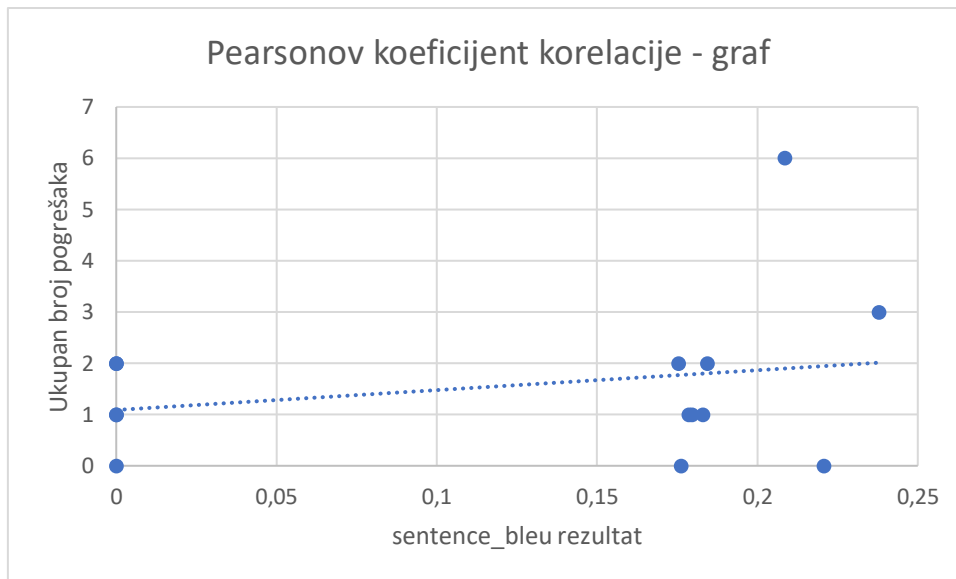
Slika 44 - Pearson - Graf - Vijesti - Bing

Koeficijent korelacije u ovom slučaju iznosi -0,001103102, iz čega je vidljivo da korelacije između dvije prikazane varijable gotovo i nema.

Google Translate

| | sentence_bleu | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------------|---------------------|-----------------------|
| 1. | 0,237795332 | 3 | 0,2446956 |
| 2. | 0 | 1 | |
| 3. | 0 | 0 | |
| 4. | 0,208599493 | 6 | |
| 5. | 0 | 1 | |
| 6. | 0,179560769 | 1 | |
| 7. | 0 | 2 | |
| 8. | 0,183050258 | 1 | |
| 9. | 0 | 2 | |
| 10. | 0,176215715 | 0 | |
| 11. | 0,184293332 | 2 | |
| 12. | 0,175410566 | 2 | |
| 13. | 0,178472607 | 1 | |
| 14. | 0,220690914 | 0 | |

Slika 45 - Pearson - Vijesti – Google



Slika 46 - Pearson - Graf - Vijesti - Google

U ovom primjeru, za razliku od slučaja s Bing prevoditeljem, značajnija je korelacija, i to pozitivna, a iznosi 0,2446956.

6.2. Pearsonov koeficijent korelacije: chrF++ metrika

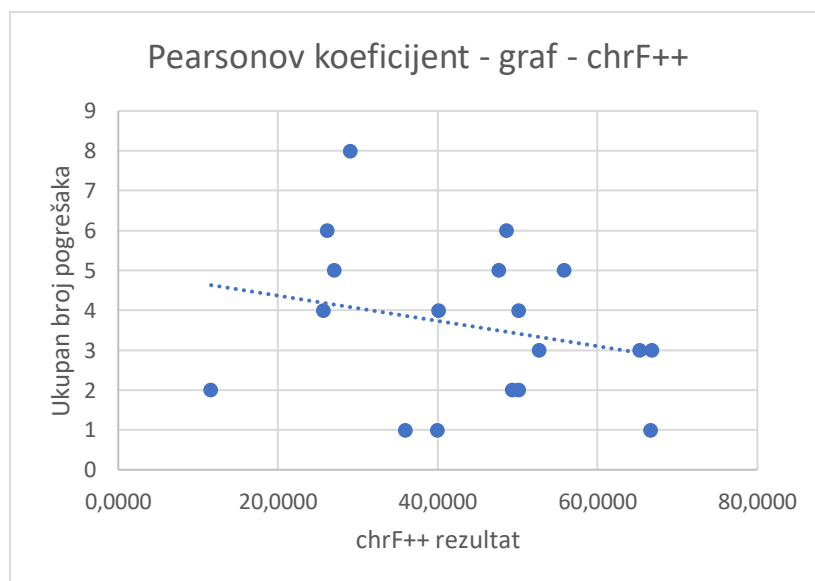
U sljedećim potpoglavljima prikazani su Pearsonovi koeficijenti za dvije varijable – rezultate chrF++ metrike na bazi rečenice i ukupnog broja pogrešaka na razini rečenice.

6.2.1. Rezultati: Vrsta teksta – Recept

Bing Microsoft Translator

| | chrF++ | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------|---------------------|-----------------------|
| 1. | 50,1095 | 4 | -0,250705047 |
| 2. | 52,6714 | 3 | |
| 3. | 39,9258 | 1 | |
| 4. | 50,177 | 2 | |
| 5. | 35,9446 | 1 | |
| 6. | 49,344 | 2 | |
| 7. | 29,0763 | 8 | |
| 8. | 48,5883 | 6 | |
| 9. | 26,1776 | 6 | |
| 10. | 47,684 | 5 | |
| 11. | 40,1437 | 4 | |
| 12. | 11,6149 | 2 | |
| 13. | 65,3154 | 3 | |
| 14. | 25,6858 | 4 | |
| 15. | 27,0699 | 5 | |
| 16. | 55,7961 | 5 | |
| 17. | 66,8136 | 3 | |
| 18. | 66,6252 | 1 | |

Slika 47 - Pearsonov koeficijent - Recept - Bing



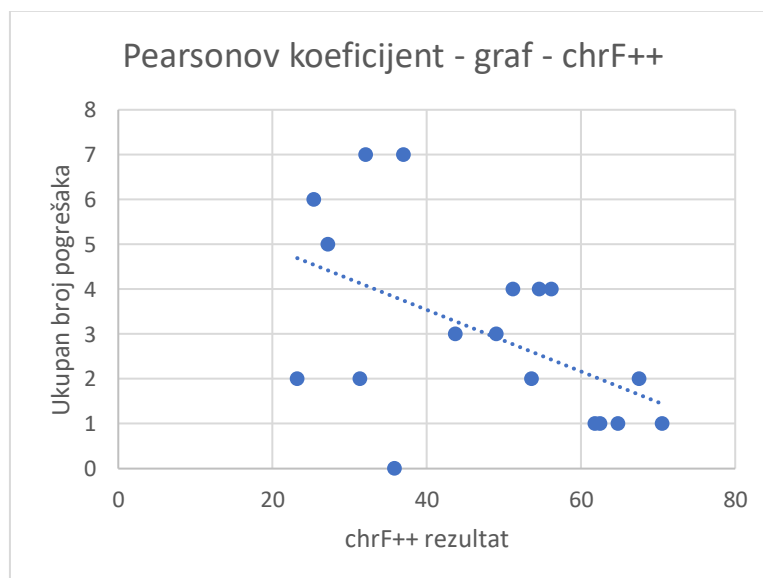
Slika 48 - Pearsonov koeficijent - Graf - Recept – Bing

U ovom slučaju chrF++ metrika daje negativan koeficijent korelacije u usporedbi s rezultatima korelacije BLEU metrike, gdje je korelacija pozitivna, ali gotovo da je i nema.

Google Translate

| | chrF++ | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------|---------------------|-----------------------|
| 1. | 67,474 | 2 | -0,500425376 |
| 2. | 53,5844 | 2 | |
| 3. | 31,2758 | 2 | |
| 4. | 61,7625 | 1 | |
| 5. | 36,9894 | 7 | |
| 6. | 49,0164 | 3 | |
| 7. | 32,0202 | 7 | |
| 8. | 62,4525 | 1 | |
| 9. | 25,3327 | 6 | |
| 10. | 51,1364 | 4 | |
| 11. | 43,7148 | 3 | |
| 12. | 35,7613 | 0 | |
| 13. | 64,7448 | 1 | |
| 14. | 23,2056 | 2 | |
| 15. | 27,1713 | 5 | |
| 16. | 56,1689 | 4 | |
| 17. | 54,5248 | 4 | |
| 18. | 70,5275 | 1 | |

Slika 49 - Pearsonov koeficijent - Recept – Google



Slika 50 - Pearsonov koeficijent - Graf - Recept - Google

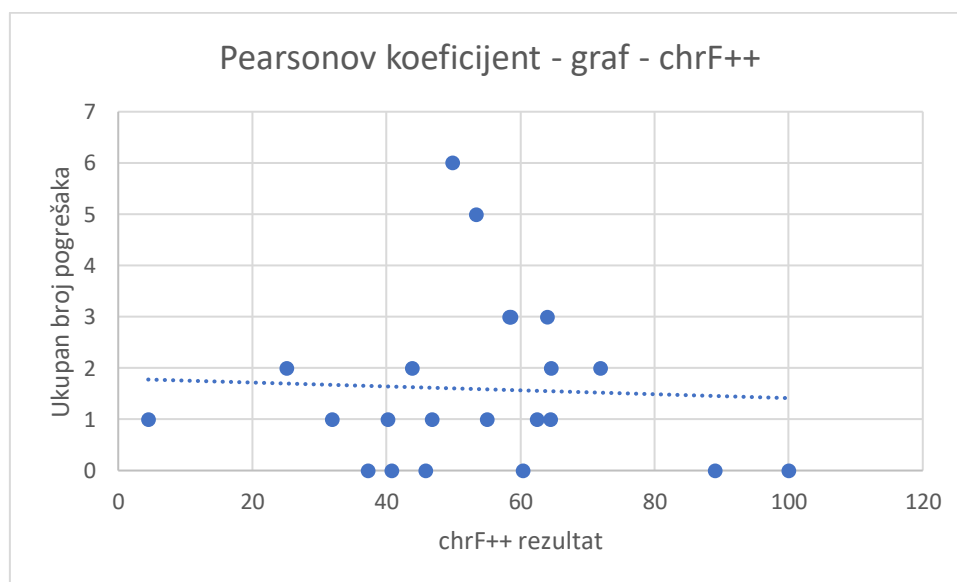
Usporedimo li rezultat korelacije Bing i Google prevoditelja za prijevod teksta „Recept“, ali i rezultate BLEU i chrF++ metrike za istu kategoriju, vidljivo je kako temeljem chrF++ metrike rezultat značajno različit – negativna korelacija koja iznosi -0,500425376.

6.2.2. Rezultati: Vrsta teksta – Upute

Bing Microsoft Translator

| | chrF++ | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------|---------------------|-----------------------|
| 1. | 100 | 0 | -0,047815209 |
| 2. | 62,4937 | 1 | |
| 3. | 64,4933 | 1 | |
| 4. | 63,9962 | 3 | |
| 5. | 55,0468 | 1 | |
| 6. | 40,1878 | 1 | |
| 7. | 37,2871 | 0 | |
| 8. | 43,8519 | 2 | |
| 9. | 64,5895 | 2 | |
| 10. | 53,4415 | 5 | |
| 11. | 25,085 | 2 | |
| 12. | 31,8757 | 1 | |
| 13. | 58,3273 | 3 | |
| 14. | 46,8008 | 1 | |
| 15. | 45,8329 | 0 | |
| 16. | 58,5807 | 3 | |
| 17. | 89,0377 | 0 | |
| 18. | 71,96 | 2 | |
| 19. | 4,4681 | 1 | |
| 20. | 60,3563 | 0 | |
| 21. | 49,8702 | 6 | |
| 22. | 40,8155 | 0 | |

Slika 51 - Pearsonov koeficijent - Upute – Bing



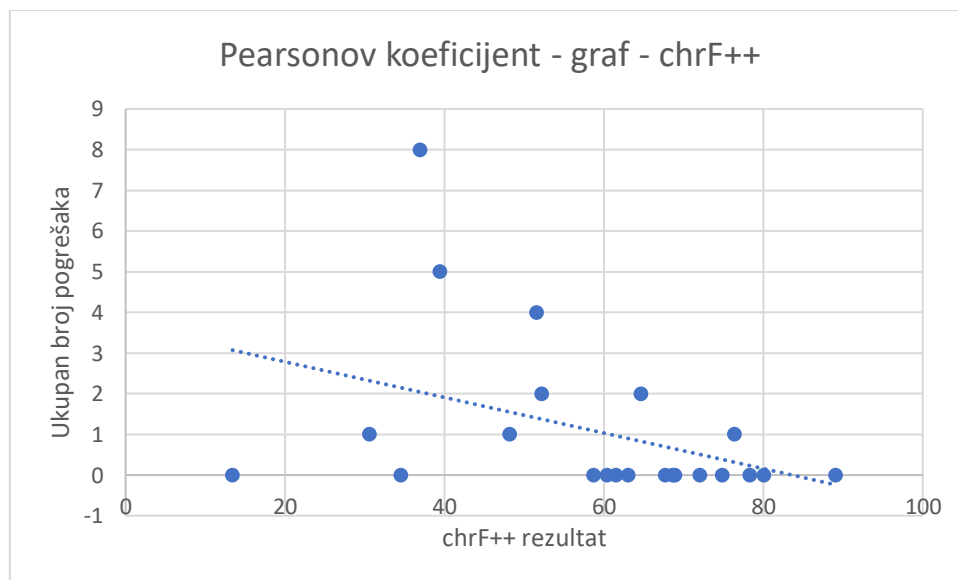
Slika 52 - Pearsonov koeficijent - Graf - Upute – Bing

Pearsonov koeficijent korelacije iznosi -0,047815209.

Google Translate

| | chrF++ | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------|---------------------|-----------------------|
| 1. | 30,5623 | 1 | -0,395449435 |
| 2. | 52,1779 | 2 | |
| 3. | 80,0722 | 0 | |
| 4. | 68,6104 | 0 | |
| 5. | 68,8478 | 0 | |
| 6. | 61,4907 | 0 | |
| 7. | 67,6736 | 0 | |
| 8. | 58,6595 | 0 | |
| 9. | 64,6221 | 2 | |
| 10. | 51,5391 | 4 | |
| 11. | 74,8622 | 0 | |
| 12. | 34,5453 | 0 | |
| 13. | 78,2641 | 0 | |
| 14. | 63,0411 | 0 | |
| 15. | 48,1447 | 1 | |
| 16. | 76,3179 | 1 | |
| 17. | 89,0377 | 0 | |
| 18. | 71,9892 | 0 | |
| 19. | 13,4187 | 0 | |
| 20. | 60,3563 | 0 | |
| 21. | 36,8777 | 8 | |
| 22. | 39,4253 | 5 | |

Slika 53 - Pearsonov koeficijent - Upute – Google



Slika 54 - Pearsonov koeficijent - Graf - Upute – Google

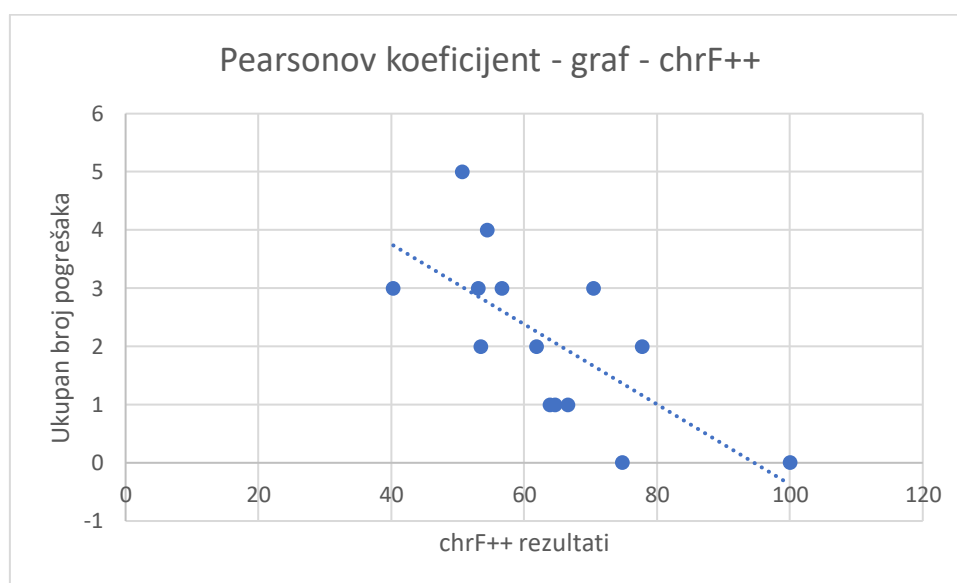
Također je i u ovom slučaju za Google prevoditelj značajna razlika u usporedbi s rezultatima koji uključuju BLEU metriku – Pearsonov koeficijent korelacije iznosi -0,395449435.

6.2.3. Rezultati: Vrsta teksta – Vijesti

Bing Microsoft Translator

| | chrF++ | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------|---------------------|-----------------------|
| 1. | 53,4918 | 2 | -0,687208489 |
| 2. | 63,8517 | 1 | |
| 3. | 66,5475 | 1 | |
| 4. | 70,4434 | 3 | |
| 5. | 50,6629 | 5 | |
| 6. | 77,7645 | 2 | |
| 7. | 56,6806 | 3 | |
| 8. | 54,4086 | 4 | |
| 9. | 53,0433 | 3 | |
| 10. | 100 | 0 | |
| 11. | 74,7717 | 0 | |
| 12. | 40,2939 | 3 | |
| 13. | 64,6816 | 1 | |
| 14. | 61,8511 | 2 | |

Slika 55 - Pearsonov koeficijent - Vijesti – Bing



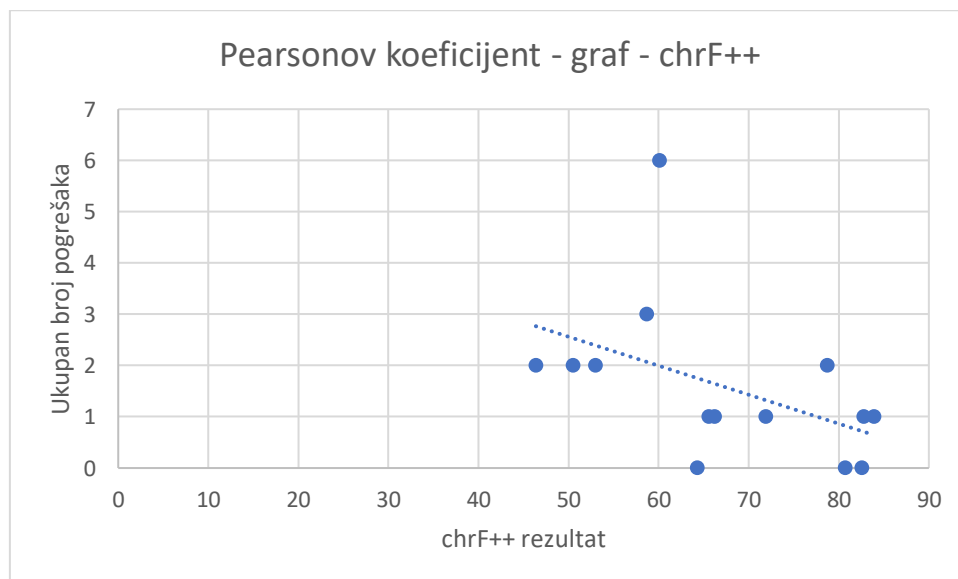
Slika 56 - Pearsonov koeficijent - Graf - Vijesti - Bing

Primjetna je razlika između rezultata BLEU i chrF++ metrike – u ovom slučaju Pearsonov koeficijent korelacije iznosi -0,687208489, što je značajan rezultat.

Google Translate

| | chrF++ | ukupan broj grešaka | Pearsons coeff. corr. |
|-----|---------|---------------------|-----------------------|
| 1. | 58,6231 | 3 | -0,467133403 |
| 2. | 65,5697 | 1 | |
| 3. | 64,2743 | 0 | |
| 4. | 60,0883 | 6 | |
| 5. | 83,9179 | 1 | |
| 6. | 82,7294 | 1 | |
| 7. | 50,4534 | 2 | |
| 8. | 71,8779 | 1 | |
| 9. | 46,3702 | 2 | |
| 10. | 82,5395 | 0 | |
| 11. | 78,7379 | 2 | |
| 12. | 52,9864 | 2 | |
| 13. | 66,1576 | 1 | |
| 14. | 80,6846 | 0 | |

Slika 57- Pearsonov koeficijent - Vijesti – Google



Slika 58 - Pearsonov koeficijent - Graf - Vijesti - Google

Pearsonov koeficijent korelacije u ovom slučaju iznosi -0,467133403.

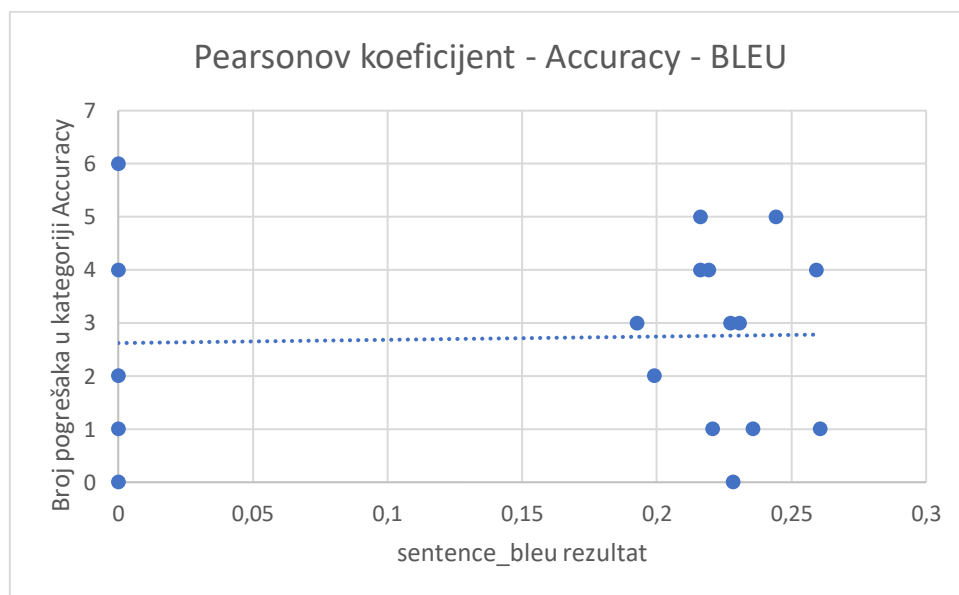
6.3. Rezultati: Accuracy i Mistranslation – BLEU metrika

Slijedi prikaz Pearsonovog koeficijenta korelacije za dvije varijable – rezultat BLEU metrike na razini rečenice i broja pogrešaka koje pripadaju kategoriji *Accuracy*, a zatim i podkategoriji *Mistranslation*.

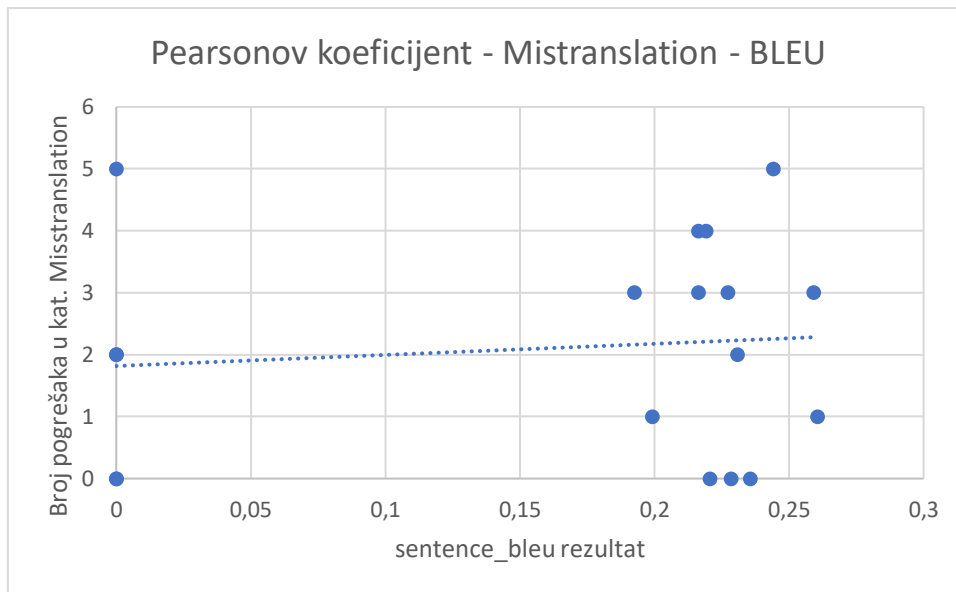
6.3.1. Accuracy i Mistranslation – tekst „Recept“

Bing Microsoft Translator: Rezultat Pearsonovog koeficijenta korelacije za varijable BLEU metrike na razini rečenice te broja pogrešaka u kategoriji *Accuracy* za svaku od rečenica iznosi 0,035868176, a za podkategoriju *Mistranslation* iznosi 0,10873846.

Prikaz grafova:



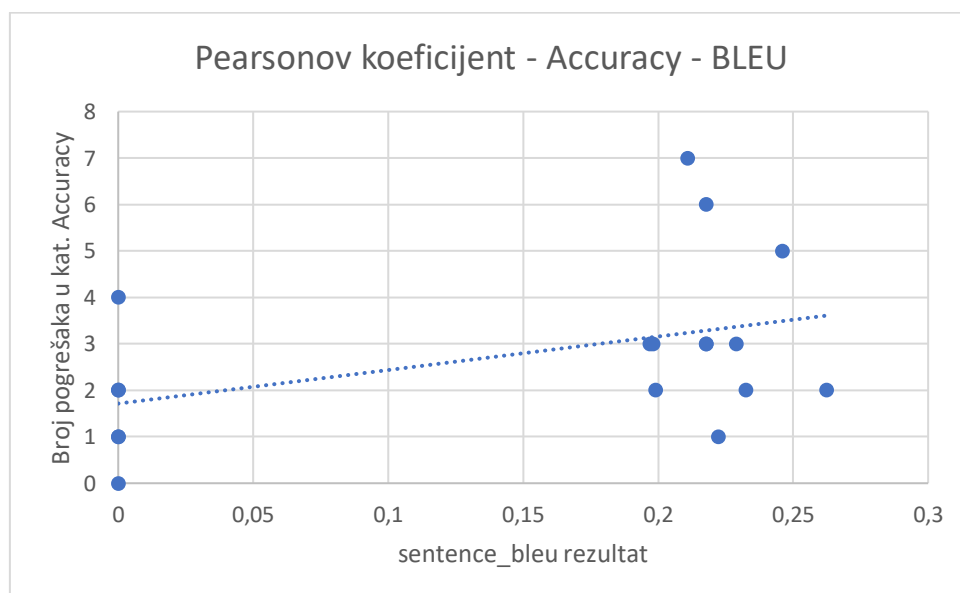
Slika 59 - Pearsonov koeficijent – Recept – Bing - Accuracy – BLEU



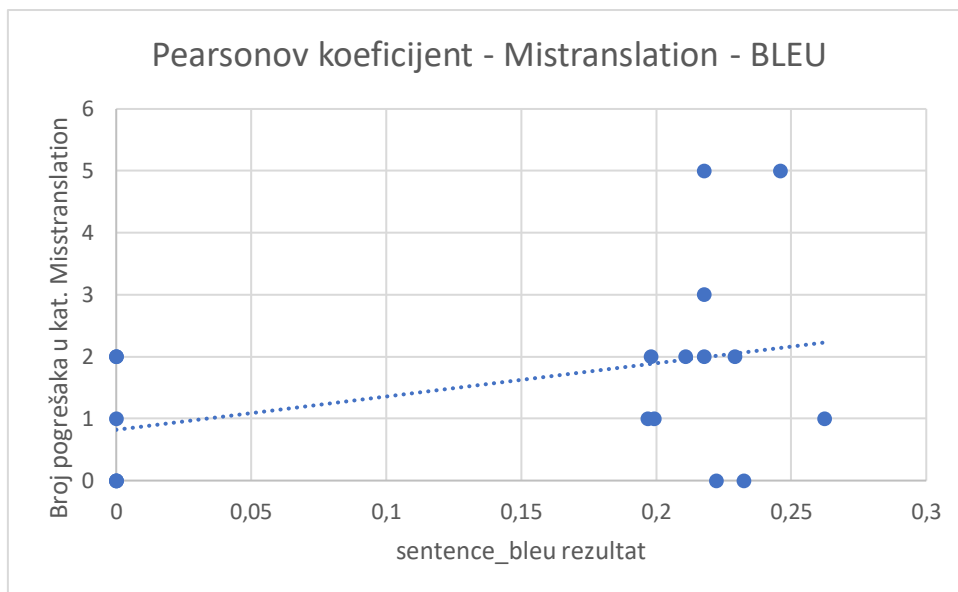
Slika 60 - Pearsonov koeficijent – Recept – Bing- Mistranslation - BLEU

Google Translate: Rezultati korelacija za varijable rezultata BLEU metrike na razini rečenice i pogrešaka kategorije *Accuracy* na razini rečenice iznosi 0,434188427, a za kategoriju *Mistranslation* iznosi 0,378038652.

Prikaz grafova:



Slika 61 - Pearsonov koeficijent – Recept – Google - Accuracy - BLEU



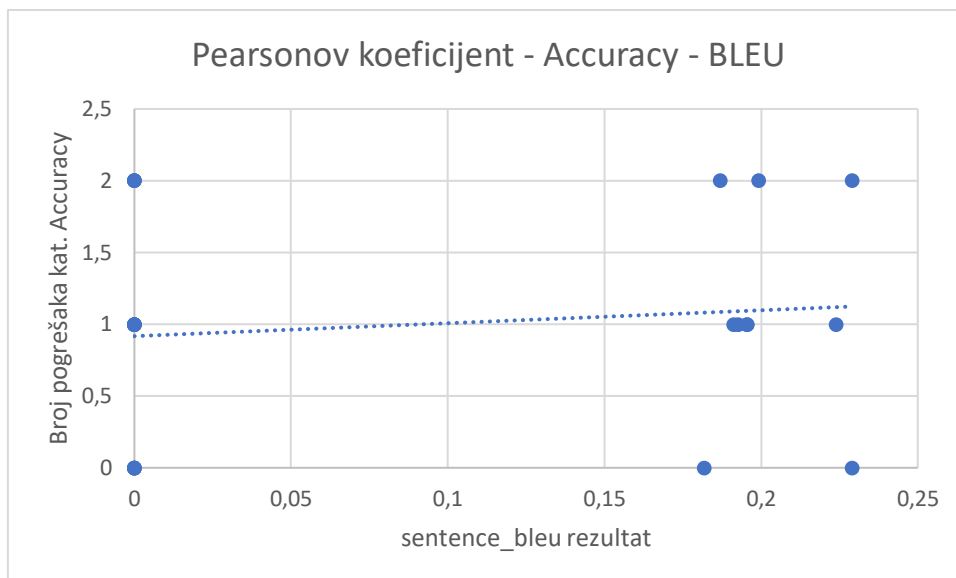
Slika 62 - Pearsonov koeficijent - Recept – Google - Mistranslation - BLEU

6.3.2.. Accuracy i Mistranslation – tekst „Upute“

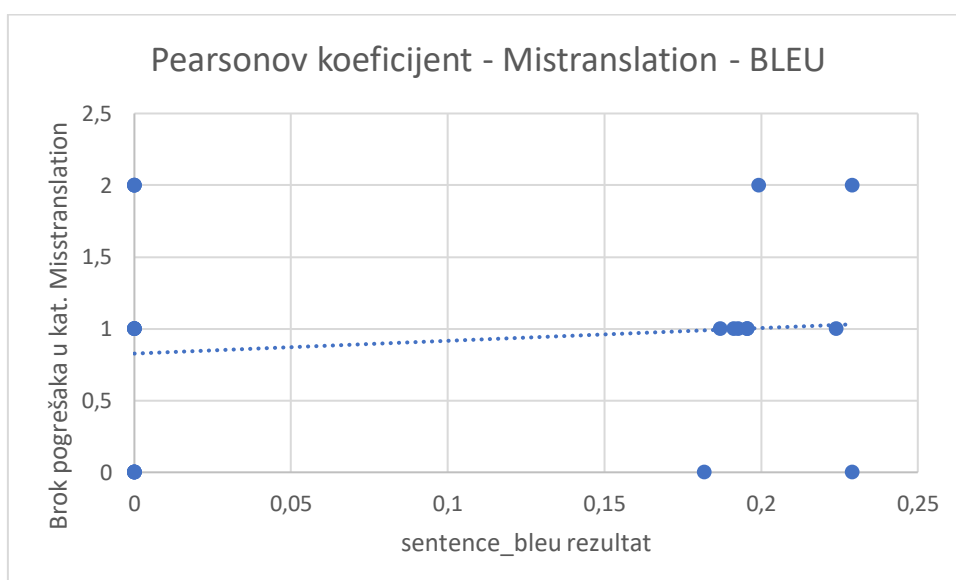
Bing Microsoft Translator: Rezultati Pearsonovog koeficijenta korelacije za varijable rezultata BLEU metrike na razini rečenice i broja pogrešaka koje pripadaju kategoriji *Accuracy* iznosi 0,123868397, a za podkategoriju *Mistranslation* iznosi 0,123076065.

Budući da skoro sve pogreške iz kategorije *Accuracy* pripadaju i podkategoriji *Mistranslation*, rezultat je gotovo izjednačen – blaga pozitivna korelacija.

Prikaz grafova:



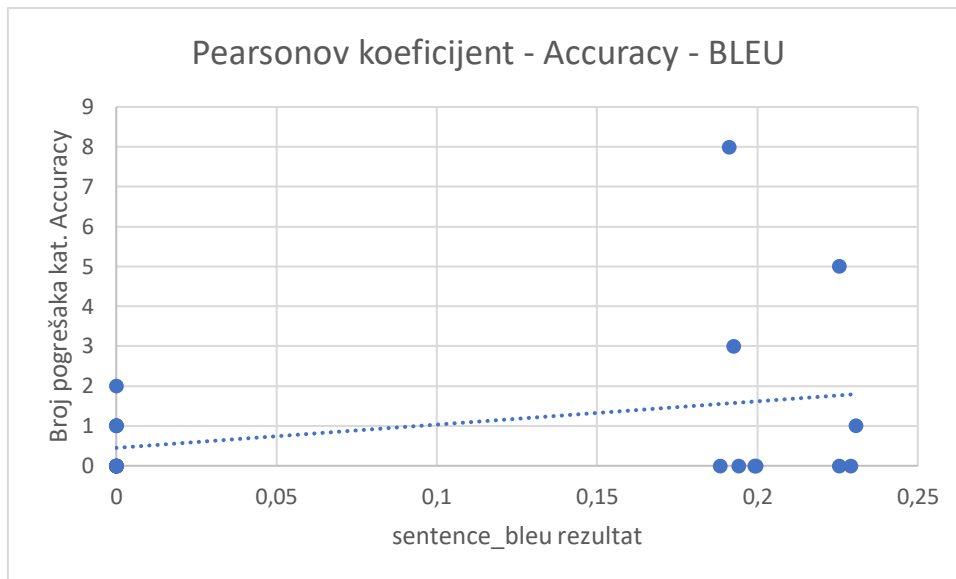
Slika 63 - Pearsonov koeficijent - Upute - Bing - Accuracy – BLEU



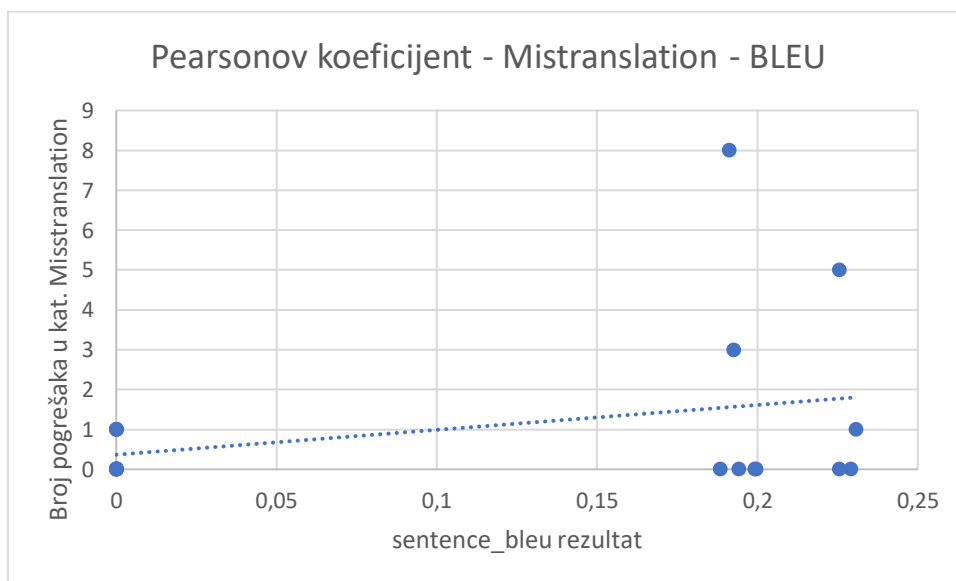
Slika 64 - Pearsonov koeficijent - Upute - Bing - Mistranslation - BLEU

Google Translate: Također su za izračun korelacije korištene varijable rezultata BLEU metrike na razini rečenice te ukupan broj pogrešaka u kategoriji *Accuracy*, što je dalo rezultat korelacije 0,310916781. Za varijablu *Mistranslation* rezultat je 0,3341109.

Prikaz grafova:



Slika 65 - Pearsonov koeficijent - Upute - Google - Accuracy – BLEU

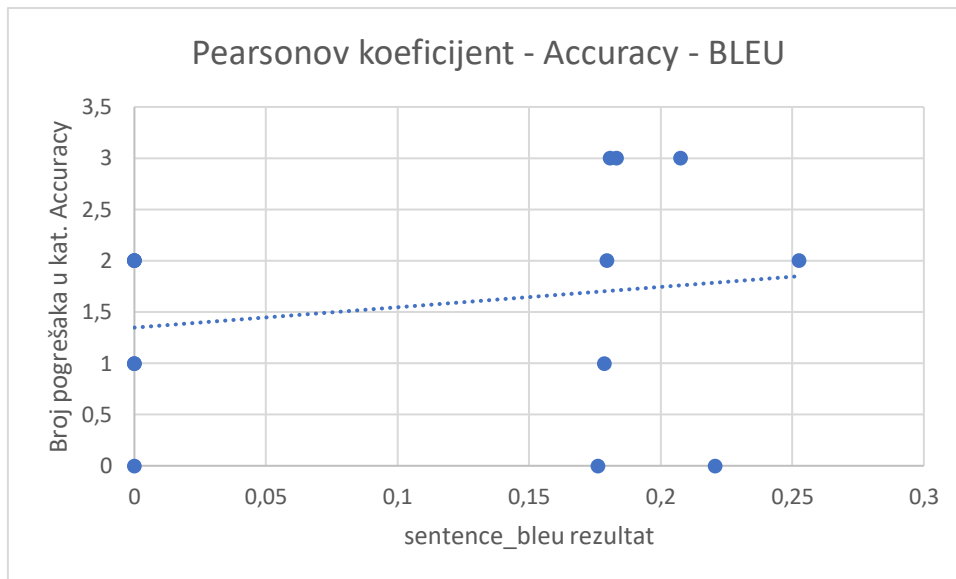


Slika 66 - Pearsonov koeficijent - Upute - Google - Mistranslation - BLEU

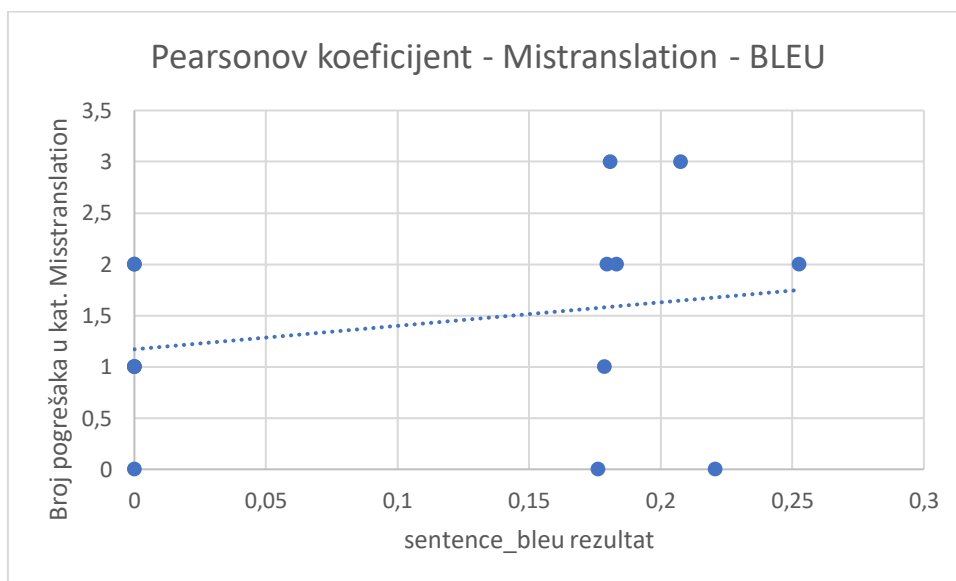
6.3.3. Accuracy i Mistranslation – tekst „Vijesti“

Bing Microsoft Translator: Rezultati Pearsonove korelacije za varijable rezultata BLEU metrike na razini rečenice i broja pogrešaka kategorije *Accuracy* iznosi 0,188233245, dok za broj pogrešaka podkategorije *Mistranslation* iznosi 0,232864749.

Prikaz grafova:



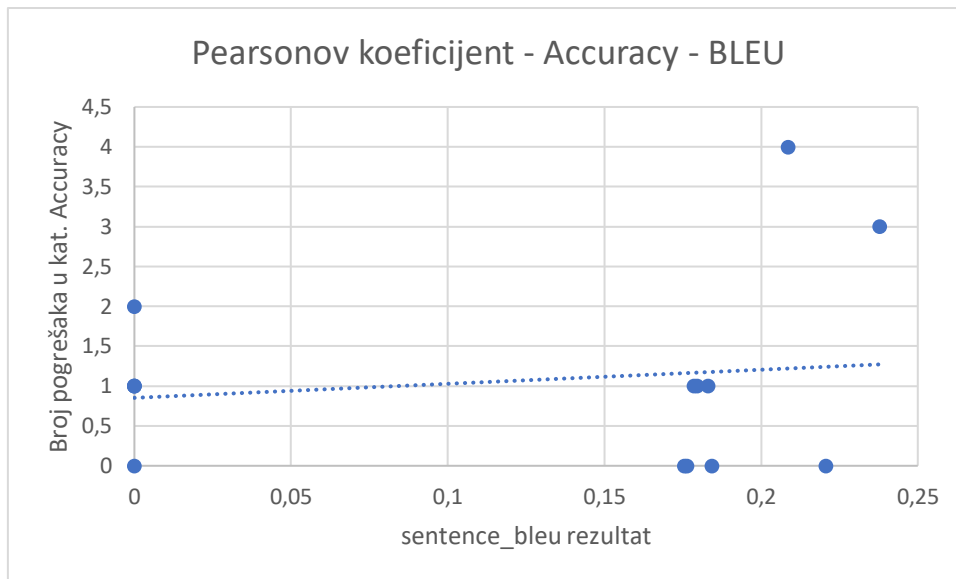
Slika 67 - Pearsonov koeficijent - Vijesti - Bing - Accuracy – BLEU



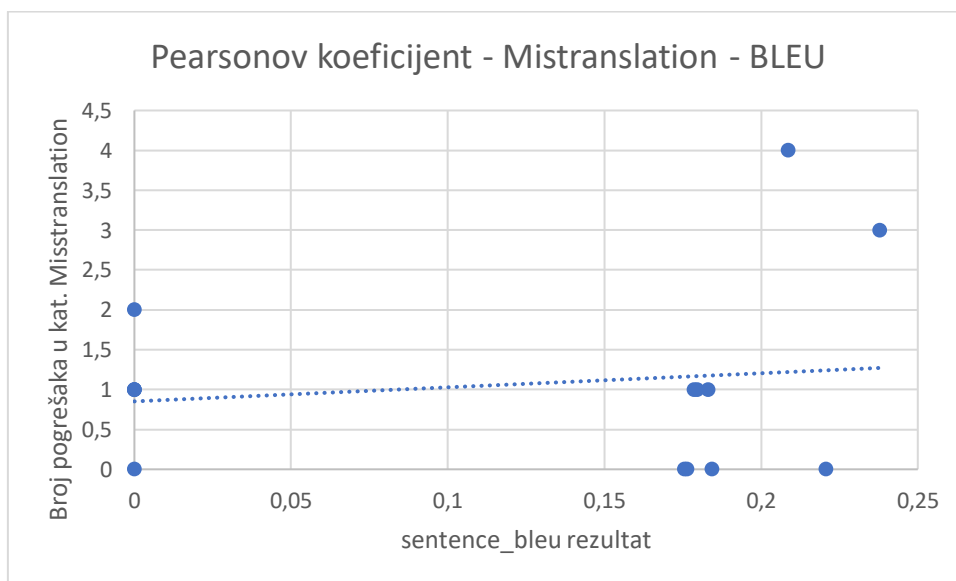
Slika 68 - Pearsonov koeficijent - Vijesti - Bing - Mistranslation - BLEU

Google Translate: Pearsonov koeficijent korelacije za varijable rezultata BLEU metrike na razini rečenice te broja pogrešaka u kategoriji *Accuracy* iznosi 0,143225037, a budući da je jednak broj pogrešaka u obje kategorije, rezultat za kategoriju *Mistranslation* je identičan (0,143225037).

Prikaz grafova:



Slika 69 - Pearsonov koeficijent - Vijesti - Google - Accuracy – BLEU



Slika 70 - Pearsonov koeficijent - Vijesti - Google - Mistranslation - BLEU

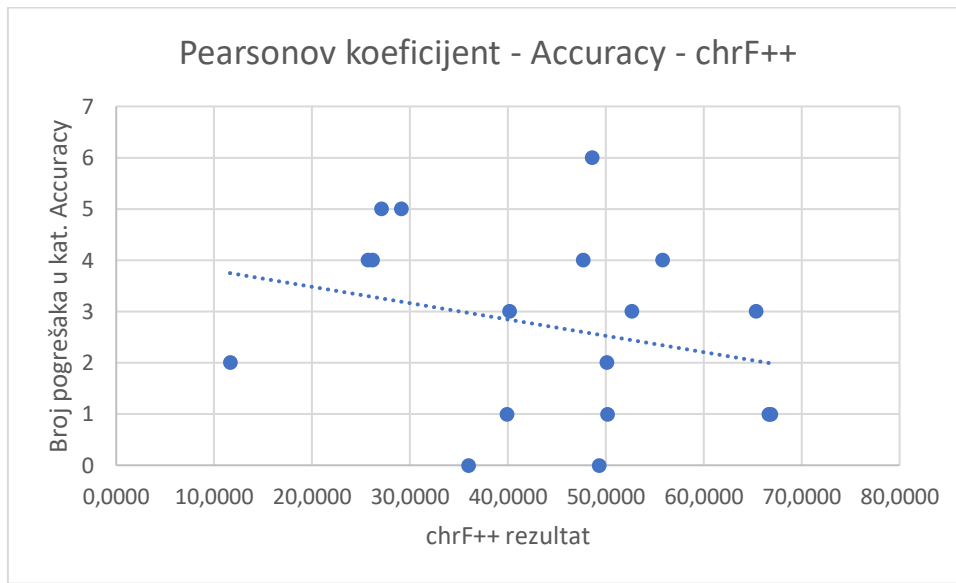
6.4. Accuracy i Mistranslation – chrF++ metrika

Slijedi prikaz rezultata Pearsonove korelacije za varijable koje obuhvaćaju rezultate chrF++ metrike na razini rečenice te broj pogrešaka koji pripadaju kategoriji *Accuracy*, a zatim i broj pogrešaka koji pripadaju podkategoriji *Mistranslation*.

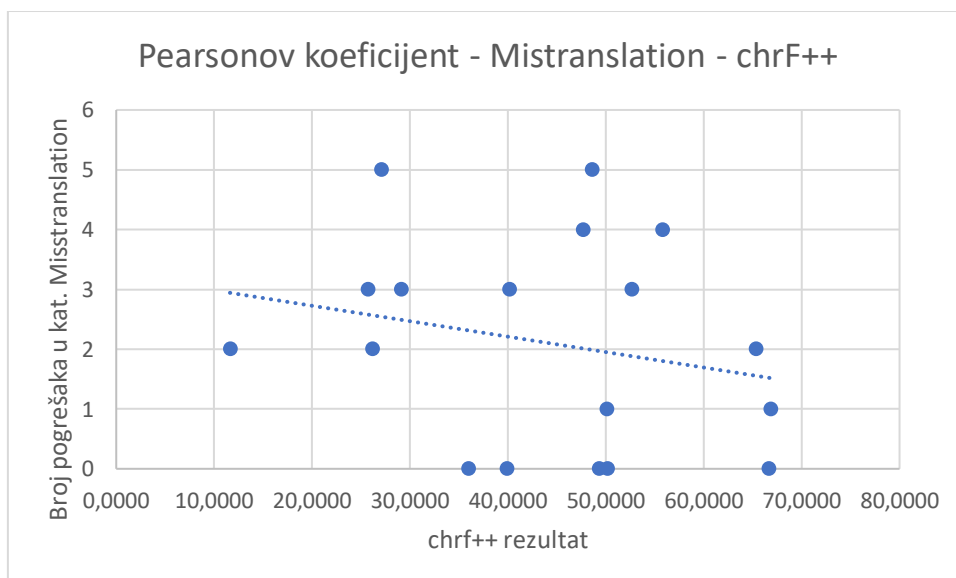
6.4.1. Accuracy i Mistranslation – tekst „Recept“

Bing Microsoft Translator: Rezultat Pearsonove korelacije za varijable chrF++ rezultata na razini rečenice te broja pogrešaka kategorije *Accuracy* iznosi -0,27523758, a za podkategoriju *Mistranslation* iznosi -0,231102907.

Prikaz grafova:



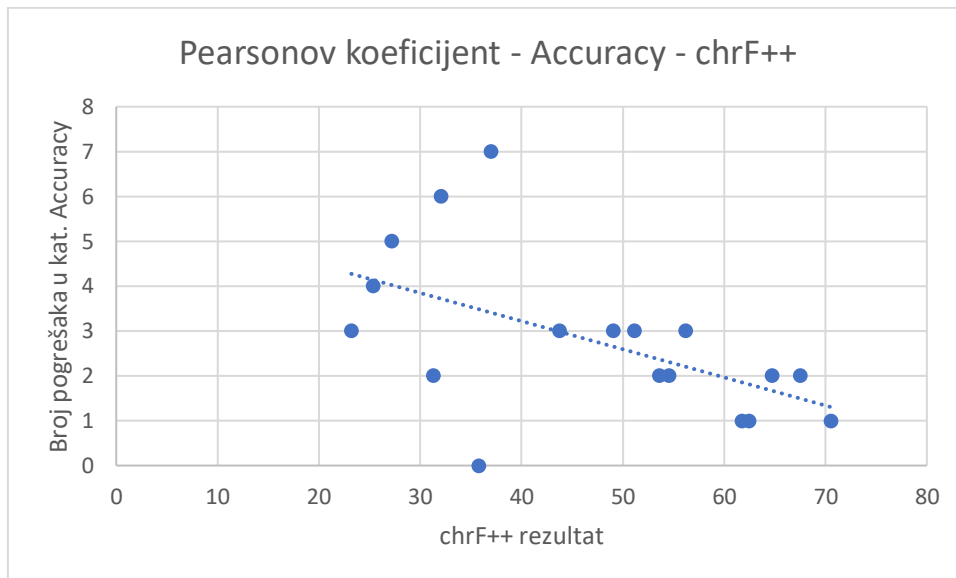
Slika 71 - Pearsonov koeficijent - Recept - Bing - Accuracy - chrF++



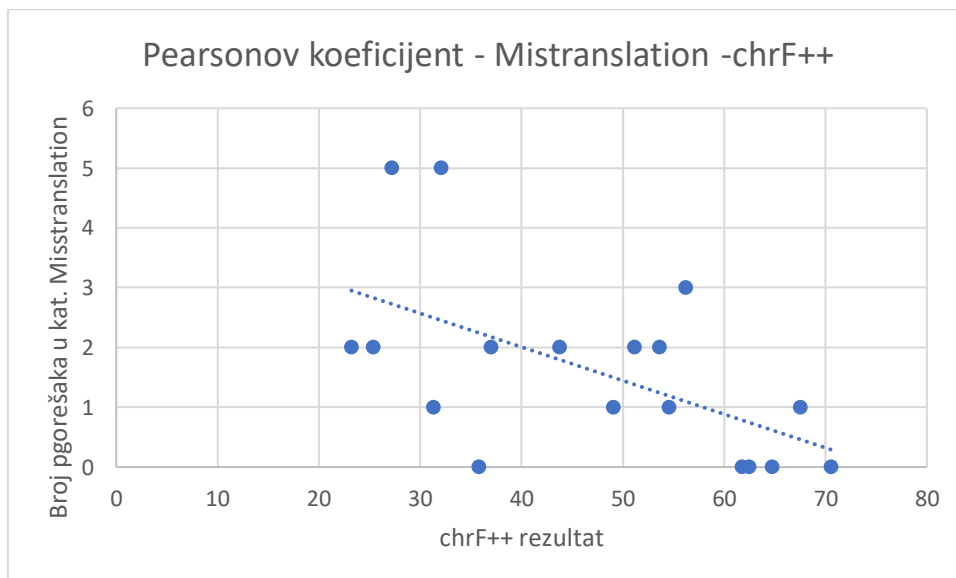
Slika 72 - Pearsonov koeficijent - Recept - Bing - Mistranslation - chrF++

Google Translate: Pearsonov koeficijent korelacije za prijevod dobiven pomoću Google Translate-a, a koji obuhvaća varijable chrF++ metrike na razini rečenice te pogrešaka u kategoriji *Accuracy* iznosi -0,540417384, a za podkategoriju *Mistranslation* -0,566145854, za što možemo reći da je značajnija negativna korelacija.

Prikaz grafova:



Slika 73 - Pearsonov koeficijent - Recept - Google - Accuracy - chrF++

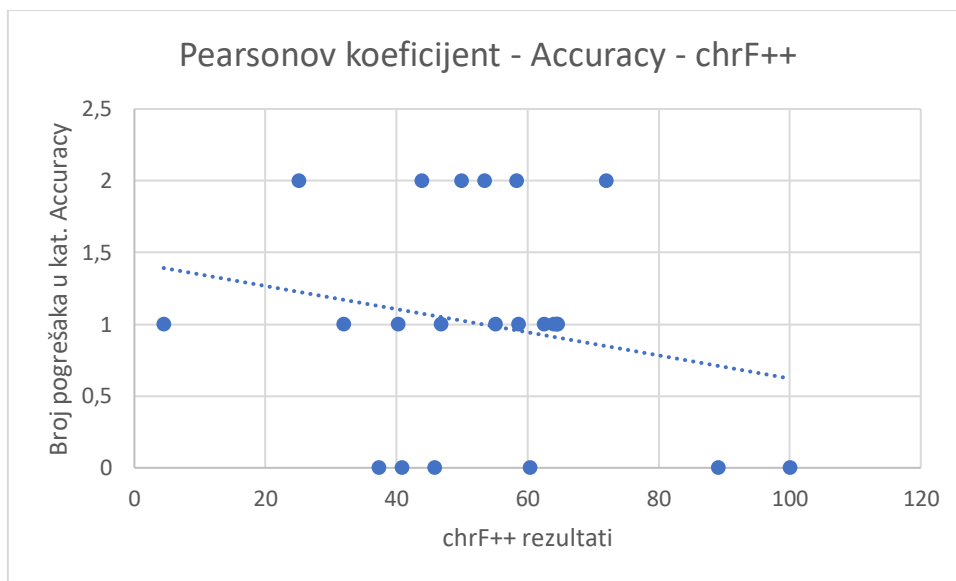


Slika 74 - Pearsonov koeficijent - Recept - Google - Mistranslation - chrF++

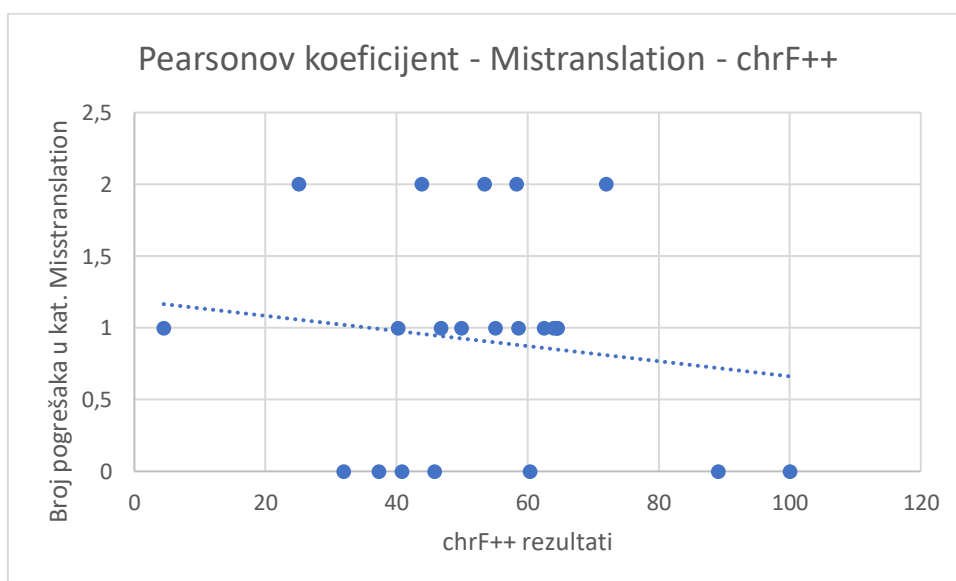
6.4.2. Accuracy i Mistranslation – tekst „Upute“

Bing Microsoft Translator: Rezultat Pearsonove korelacije, dobiven na temelju prijevoda teksta „Upute“ pomoću Bing prevoditelja, iznosi -0,21802527 za varijable rezultat chrF++ metrike na razini rečenice te kategorije *Accuracy*. Za kategoriju *Mistranslation* Pearsonova korelacija iznosi -0,143750955.

Prikaz grafova:



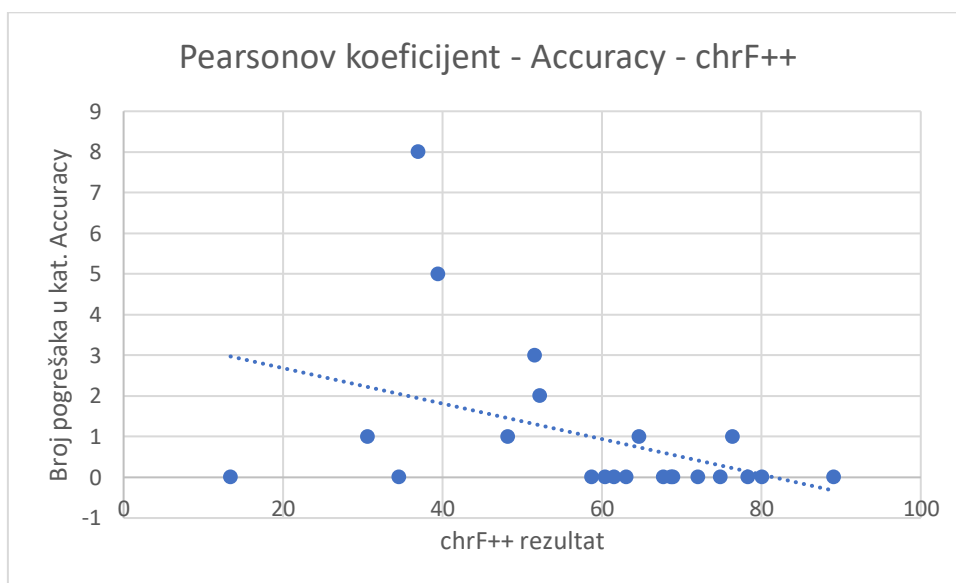
Slika 75 - Pearsonov koeficijent - Upute - Bing - Accuracy - chrF++



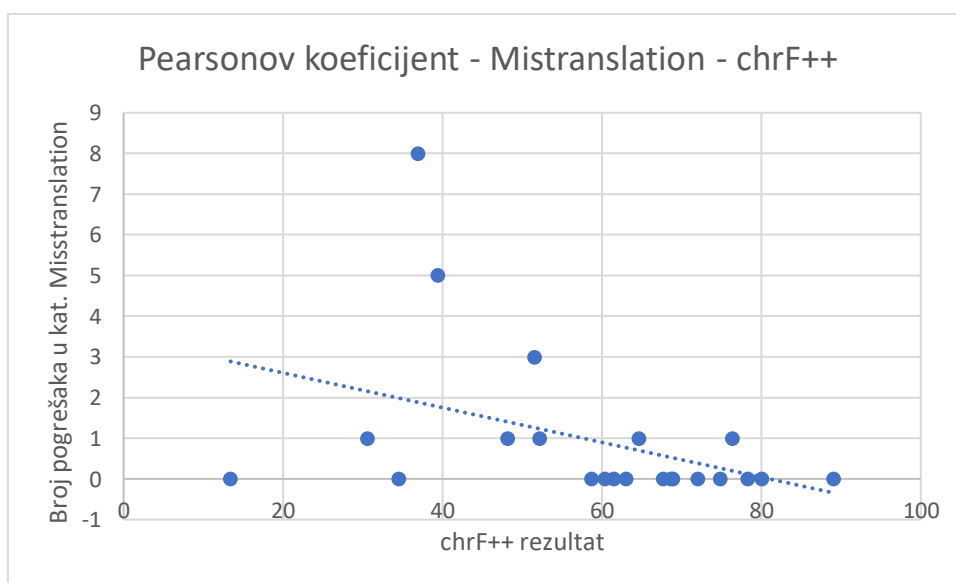
Slika 76 - Pearsonov koeficijent - Upute - Bing - Mistranslation - chrF++

Google Translate: Za varijable rezultata chrF++ metrike na razini rečenice te broja pogrešaka u kategoriji *Accuracy*, prijevod dobiven pomoću prevoditelja Google Translate dao je kao rezultat Pearsonove korelacije iznos -0,407437047, a za podkategoriju *Mistranslation* - 0,401684345, što je poprilično jednak rezultat.

Prikaz grafova:



Slika 77 - Pearsonov koeficijent - Upute - Google - Accuracy - chrF++

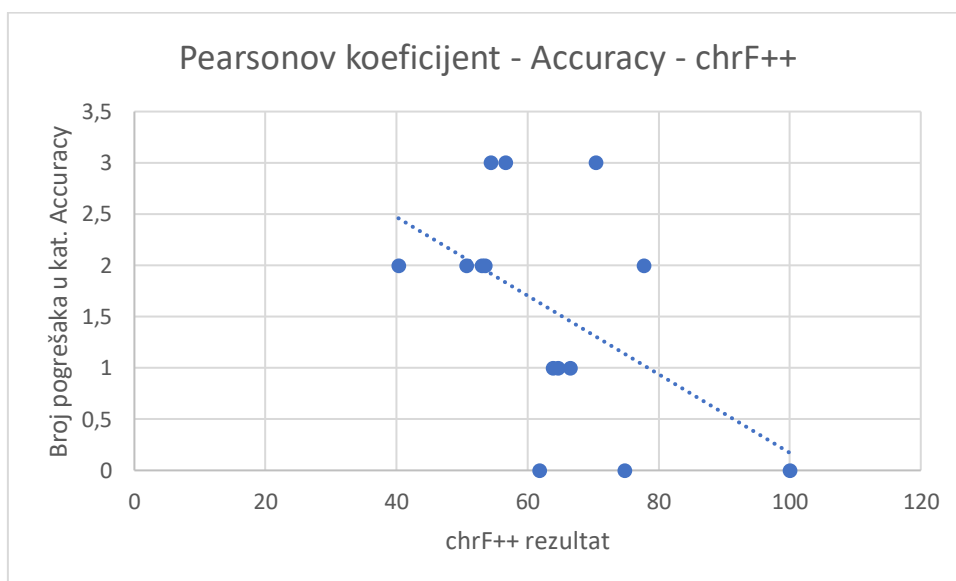


Slika 78 - Pearsonov koeficijent - Upute - Google - Mistranslation - chrF++

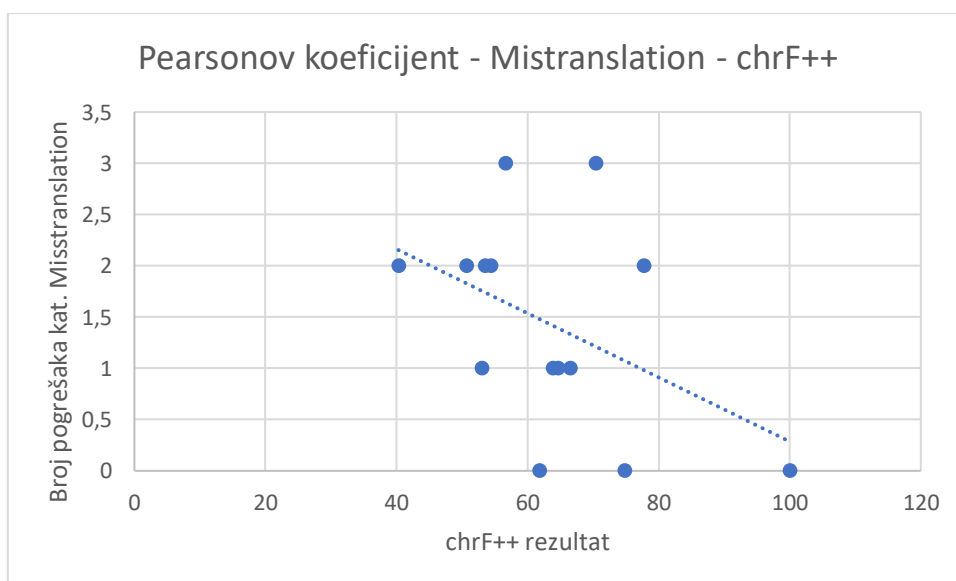
6.4.3. Accuracy i Mistranlation – tekst „Vijesti“

Bing Microsoft Translator: Za vrstu teksta „Vijesti“, prijevod dobiven pomoću online prevoditelja Bing, rezultat Pearsonove korelacije za varijable rezultata chrF++ metrike na razini rečenice i broja pogrešaka u kategoriji *Accuracy* iznosi -0,513128118, a za podkategoriju *Mistranlation* -0,448969464.

Prikaz grafova:



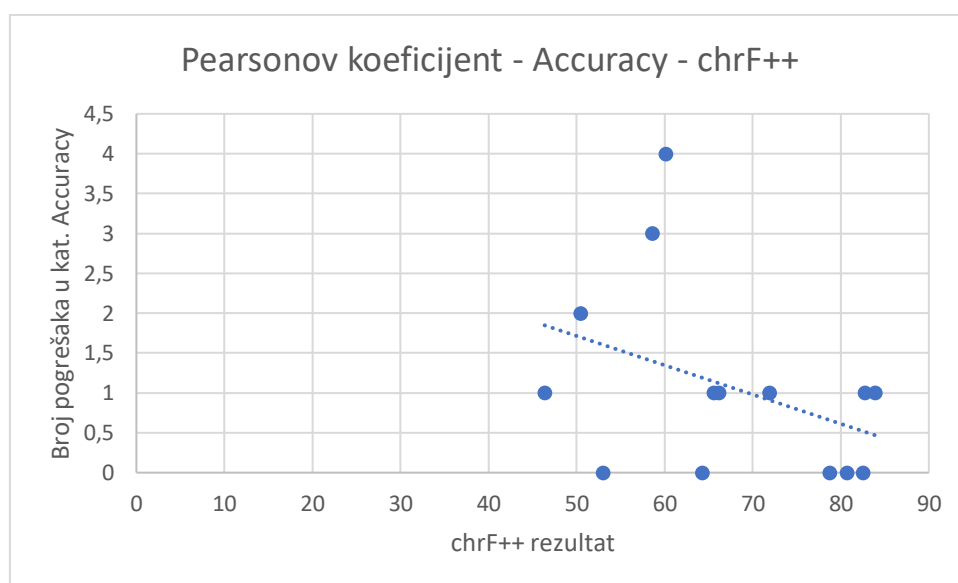
Slika 79 - Pearsonov koeficijent - Vijesti - Bing - Accuracy - chrF++



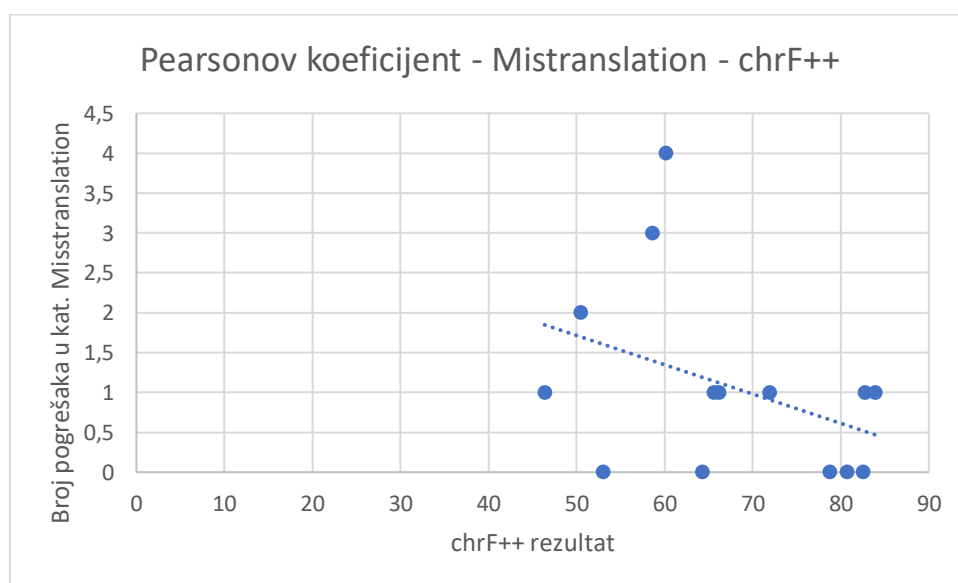
Slika 80 - Pearsonov koeficijent - Vijesti - Bing - Mistranlation - chrF++

Google Translate: Pearsonov koeficijent korelacije, za prijevod dobiven pomoću online prevoditelja Google Translate, iznosi -0,390926992 za varijable rezultata chrF++ metrike na razini rečenice te broja pogrešaka kategorije *Accuracy*, a za podkategoriju *Mistranslation* iznosi -0,390926992.

Prikaz grafova:



Slika 81 - Pearsonov koeficijent - Vijesti - Google - Accuracy - chrF++



Slika 82 - Pearsonov koeficijent - Vijesti - Google - Mistranslation - chrF++

6.4.4. Rezultati: Accuracy i Mistranslation – tablica

Tablica 4 prikazuje rezultate Pearsonove korelacije za svaku vrstu teksta za kategoriju *Accuracy* (u tablici 4 označena slovom "A"), a tablica 5 prikazuje rezultate Pearsonove korelacije za podkategoriju *Mistranslation*.

U ovim tablicama redak pod nazivom „BLEU metrika“ označava da je kao varijabla korišten rezultat dobiven pomoću BLEU metrike na razini rečenice te ukupan broj pogrešaka ovisno o kategoriji, dok redak pod nazivom „chrF++ metrika“ označava da je kao varijabla korišten rezultat dobiven pomoću chrF++ metrike na razini rečenice.

| Rezultati: | 1. tekst: Recept | | 2. tekst: Upute | | 3. tekst: Vijesti | |
|-----------------------|------------------|-------------|-----------------|-------------|-------------------|-------------|
| | Bing | Google | Bing | Google | Bing | Google |
| | A | A | A | A | A | A |
| BLEU metrika | 0,035868176 | 0,434188427 | 0,123868397 | 0,310916781 | 0,188233245 | 0,143225037 |
| chrF++ metrika | -0,27523758 | -0,54041738 | -0,21802527 | -0,40743704 | -0,51312811 | -0,39092699 |

Tablica 4 - Prikaz rezultata Pearsonove korelacije – kategorija Accuracy

| Rezultati: | 1. tekst: Recept | | 2. tekst: Upute | | 3. tekst: Vijesti | |
|-----------------------|------------------|-------------|-----------------|-------------|-------------------|-------------|
| | Bing | Google | Bing | Google | Bing | Google |
| | M | M | M | M | M | M |
| BLEU metrika | 0,10873846 | 0,378038652 | 0,123076065 | 0,3341109 | 0,232864749 | 0,143225037 |
| chrF++ metrika | -0,2311029 | -0,56614585 | -0,14375095 | -0,40168434 | -0,44896946 | -0,39092699 |

Tablica 5 - Prikaz rezultata Pearsonove korelacije - kategorija Mistranslation

Iz tablica je vidljivo kako prijevodi, evaluirani pomoću BLEU metrike imaju pozitivnu korelaciju, što znači da se povećanjem broja pogrešaka u rečenici povećava i rezultat BLEU metrike na razini rečenice. Prijevodi evaluirani pomoću chrF++ metrike imaju negativne korelacije – povećanjem broja pogrešaka na razini rečenice smanjuje se rezultat chrF++ metrike.

6.5. Pearsonov koeficijent korelacije – diskusija

| Rezultati: Pearsonov koeficijent korelacije | 1. tekst: Recept | | 2. tekst: Upute | | 3. tekst: Vijesti | |
|--|------------------|-------------|-----------------|-------------|-------------------|-------------|
| | Bing | Google | Bing | Google | Bing | Google |
| BLEU metrika | 0,001611393 | 0,382216419 | 0,346160958 | 0,212577693 | -0,0011031 | 0,2446956 |
| chrF++ metrika | -0,25070504 | -0,50042537 | -0,04781520 | -0,39544943 | -0,6872084 | -0,46713340 |

Tablica 6 - Pearsonov koeficijent korelacije - rezultati

Mnogi rezultati prikazuju pozitivnu korelaciju varijabli, posebno kod rezultata u kojima su kao varijabla korišteni rezultati BLEU metrike na razini rečenice. Kao što je već objašnjeno, pozitivna korelacija znači da se povećanjem vrijednosti jedne varijable povećava i vrijednost druge varijable.

U ovom konkretnom slučaju to bi značilo da se povećanjem broja pogrešaka na razini rečenice povećava i konačan rezultat BLEU metrike na razini rečenice. Međutim, veća količina pogrešaka u pravilu označava i lošiji prijevod, što bi značilo da bi u tom slučaju rezultati BLEU metrike, koja se koristi kao računalna evaluacija, trebali biti lošiji.

S druge strane, chrF++ metrika, koja je i u usporedbi s ljudskom evaluacijom dala bolje rezultate od BLEU metrike ima gotovo sve negativne korelacije – povećanjem vrijednosti jedne varijable smanjuje se vrijednost druge varijable. Konkretno, povećanjem broja pogrešaka na razini rečenice, chrF++ rezultat na razini rečenice je manji.

7. Zaključak

Strojno prevođenje se već godinama redovito unapređuje i poboljšava, što zasigurno dovodi do sve češće uporabe strojnih prevoditelja kod različitih „profila ljudi“ – primjerice sve od akademika pa do ljudi koji jednostavno žele razumjeti neki tekst napisan na stranome jeziku.

Za poboljšanje strojnih prevoditelja svakako je najbitnije razumijevanje same lingvistike i jezičnih pravila, pojava i osobina, a osim toga, jedan od iznimno važnih faktora je neprestano dodavanje novih riječi, odnosno ažuriranje vokabulara prevoditelja za svaki pojedini jezik.

Korištenje Google i Bing prevoditelja svakako olakšava prijevod, unatoč činjenici da oba još uvijek imaju relativno velik broj pogrešaka - posebice Bing prevoditelj. Korisnici koji se već godinama služe online prevoditeljima zasigurno su primijetili velike razlike u samoj kvaliteti prijevoda i njihovom razvoju i poboljšanju. Danas su online prevoditelji „sposobni“ točno prevesti fraze, riječi i termine koji možda čak i nisu toliko učestali ili poznati i za koje korisnik možda ni ne očekuje da će biti točno prevedeni, dok s druge strane još uvijek postoje slučajevi u kojima proces prevođenja pomoću online prevoditelja ne rezultira točnim prijevodom, iako se korisniku može činiti da se radi o nekoj banalnoj terminologiji i sl.

Za tekstove analizirane u ovome seminarskome radu, BLEU metrika u potpunosti daje prednost Bing Microsoft Translatoru, iako je razlika u rezultatima relativno mala. ChrF++ s druge strane daje prednost Google prevoditelju pa je tako u 2 od 3 slučaja Google prevoditelj bio bolji od Bing prevoditelja, iako se ni prilikom korištenja ove metrike ne može reći da je razlika u rezultatima posebno velika.

Što se ljudske evaluacije tiče, uzimajući u obzir samu popularnost Google prevoditelja, bilo je očekivano da će Google prevoditelj dati bolje rezultate jer je upravo njegova popularnost jedan od razloga za svakodnevno poboljšanje sustava prijevoda, što ujedno omogućuje i osigurava zadovoljstvo krajnjih korisnika. Međutim, Bing prevoditelj, iako možda i ne toliko korišten, nije u potpunosti podbacio te može ubrzati proces prijevoda nekih jednostavnijih vrsta tekstova ili omogućiti korisniku osnovno razumijevanje sadržaja teksta. Za stručnije tekstove koje koriste izraze specifične za određena područja, svakako bi bilo preporučljivije koristiti Google prevoditelj budući da se njegova baza podataka neprestano unapređuje i usavršava, što rezultira manjim brojem pogrešaka kao i boljom kvalitetom prijevoda – primjerice od podudaranja riječi u rodu, broju i padežu pa sve do točnog prijevoda izraza iz različitih područja.

8. Literatura

1. Brkić Bakarić, M. (2018). *Prevoditeljske tehnologije – skripta za kolegij ICT za prevoditelje*. Rijeka: Sveučilište u Rijeci.
2. Brownlee, Jason (2017). *A Gentle Introduction to Calculating the BLEU Score for Text in Python*. Dostupno na: <https://machinelearningmastery.com/calculate-bleu-score-for-text-python/> (22.2.2019.)
3. Klubička, Filip, Toralb, Antonio, Sánchez-Cartagena, Víctor M. (2017). *Fine-grained human evaluation of neural versus phrase-based machine translation*. The Prague Bulletin of Mathematical Linguistics. Dostupno na: <https://arxiv.org/pdf/1706.04389.pdf> (22.2.2019)
4. Popović, Maja (2015). *CHRF: character n-gram F-score for automatic MT evaluation*. Dostupno na: https://www.researchgate.net/publication/281677746_chrF_character_n-gram_F-score_for_automatic_MT_evaluation (20.2.2019)
5. Popović, Maja (2017). *chrF ++ : words helping character n-grams*. Dostupno na: https://www.researchgate.net/publication/319123639_chrF_words_helping_character_n-grams (20.2.2019)
6. Popović, Maja. *chrF - a tool for calculating character n-gram F score*. Dostupno na: <https://github.com/m-popovic/chrF> (20.2.2019)
7. *Guide to selecting MQM issues for the MT Evaluation Metric*. Dostupno na: <http://www.qt21.eu/downloads/annotatorsGuidelines-2014-06-11.pdf> (23.2.2019)
8. *Pearson Correlations – Quick Introduction*. Dostupno na: <https://www.spss-tutorials.com/pearson-correlation-coefficient/> (8.7.2019)
9. Dalinina,R. (31.1.2017) *Introduction to Correlation*. Dostupno na: <https://www.datascience.com/learn-data-science/fundamentals/introduction-to-correlation-python-data-science> (8.7.2019)
10. Papineni, K., Roukos, S., Ward, T.,Zhu, W.(2002) *BLEU: a Method for Automatic Evaluation of Machine Translation*. <https://www.aclweb.org/anthology/P02-1040>

POPIS SLIKA

| | |
|--|----|
| Slika 1 - Primjer funkcije sentence_bleu() | 3 |
| Slika 2 - Primjer funkcije corpus_bleu() | 3 |
| Slika 3 - izračun kumulativnog rezultata | 3 |
| Slika 4 - Prikaz teksta "Gugelhupf" | 4 |
| Slika 5 - Prikaz koda za corpus_bleu() – recept - Bing | 5 |
| Slika 6 - Rezultat corpus_bleu() - recept - Bing | 5 |
| Slika 7 - Prikaz koda za corpus_bleu() - recept - Google | 5 |
| Slika 8 - Rezultat corpus_bleu() - recept - Google | 5 |
| Slika 9 - Prikaz teksta - upute | 6 |
| Slika 10 - Prikaz koda za corpus_bleu() - upute – Bing | 7 |
| Slika 11 - Rezultat corpus_bleu() - upute - Bing | 7 |
| Slika 12 - Prikaz koda za corpus_bleu() - upute - Google | 7 |
| Slika 13 - Rezultat corpus_bleu() - upute - Google | 7 |
| Slika 14 - Prikaz teksta - Vijesti | 8 |
| Slika 15 - Prikaz koda za corpus_bleu() - vijesti - Bing | 8 |
| Slika 16 - Rezultat corpus_bleu() - vijesti - Bing | 9 |
| Slika 17 - Prikaz koda za corpus_bleu() - vijesti - Google | 9 |
| Slika 18 - Rezultat corpus_bleu() - vijesti – Google | 9 |
| Slika 19 - Prikaz MQM jezgre | 16 |
| Slika 20 - MQM - Slavic tagset | 17 |
| Slika 21 - Treat - recept – Bing 1 | 20 |
| Slika 22 - Treat - recept - Bing 2 | 20 |
| Slika 23 - Treat - recept - Bing 3 | 21 |
| Slika 24 - Treat - recept - Bing 4 | 21 |
| Slika 25 - Treat - recept - Google 1 | 22 |
| Slika 26 - Treat - recept - Google 2 | 22 |
| Slika 27 - Treat - upute - Bing 1 | 23 |
| Slika 28 - Treat - upute - Bing 2 | 23 |
| Slika 29 - Treat - upute - Google 1 | 24 |
| Slika 30 - Treat - upute - Google 2 | 24 |
| Slika 31 - Treat - vijesti - Bing 1 | 25 |
| Slika 32 - Treat - vijesti - Bing 1 | 25 |
| Slika 33 - Treat - vijesti - Google 1 | 26 |
| Slika 34 - Treat - vijesti - Google 2 | 26 |
| Slika 35 - Pearson - Recept - Bing | 28 |
| Slika 36 - Pearson - Graf - Recept - Bing | 29 |
| Slika 37 - Pearson - Recept - Google | 29 |
| Slika 38 - Pearson - Graf - Recept - Google | 30 |
| Slika 39 - Pearson – Upute – Bing | 30 |
| Slika 40 - Pearson – Graf – Upute – Bing | 31 |
| Slika 41 - Pearson – Upute – Google | 31 |
| Slika 42 - Pearson – Graf - Upute – Google | 32 |
| Slika 43 - Pearson - Vijesti – Bing | 32 |

| | |
|---|----|
| Slika 44 - Pearson - Graf - Vijesti - Bing | 33 |
| Slika 45 - Pearson - Vijesti – Google | 33 |
| Slika 46 - Pearson - Graf - Vijesti - Google..... | 34 |
| Slika 47 - Pearsonov koeficijent - Recept - Bing | 35 |
| Slika 48 - Pearsonov koeficijent - Graf - Recept – Bing | 35 |
| Slika 49 - Pearsonov koeficijent - Recept – Google | 36 |
| Slika 50 - Pearsonov koeficijent - Graf - Recept - Google..... | 36 |
| Slika 51 - Pearsonov koeficijent - Upute – Bing..... | 37 |
| Slika 52 - Pearsonov koeficijent - Graf - Upute – Bing | 37 |
| Slika 53 - Pearsonov koeficijent - Upute – Google..... | 38 |
| Slika 54 - Pearsonov koeficijent - Graf - Upute – Google..... | 38 |
| Slika 55 - Pearsonov koeficijent - Vijesti – Bing..... | 39 |
| Slika 56 - Pearsonov koeficijent - Graf - Vijesti - Bing | 39 |
| Slika 57- Pearsonov koeficijent - Vijesti – Google..... | 40 |
| Slika 58 - Pearsonov koeficijent - Graf - Vijesti - Google | 40 |
| Slika 59 - Pearsonov koeficijent – Recept – Bing - Accuracy – BLEU..... | 41 |
| Slika 60 - Pearsonov koeficijent – Recept – Bing- Mistranslation - BLEU | 42 |
| Slika 61 - Pearsonov koeficijent – Recept – Google - Accuracy - BLEU..... | 42 |
| Slika 62 - Pearsonov koeficijent - Recept – Google - Mistranslation - BLEU..... | 43 |
| Slika 63 - Pearsonov koeficijent - Upute - Bing - Accuracy – BLEU | 44 |
| Slika 64 - Pearsonov koeficijent - Upute - Bing - Mistranslation - BLEU..... | 44 |
| Slika 65 - Pearsonov koeficijent - Upute - Google - Accuracy – BLEU..... | 45 |
| Slika 66 - Pearsonov koeficijent - Upute - Google - Mistranslation - BLEU..... | 45 |
| Slika 67 - Pearsonov koeficijent - Vijesti - Bing - Accuracy – BLEU..... | 46 |
| Slika 68 - Pearsonov koeficijent - Vijesti - Bing - Mistranslation - BLEU..... | 46 |
| Slika 69 - Pearsonov koeficijent - Vijesti - Google - Accuracy – BLEU..... | 47 |
| Slika 70 - Pearsonov koeficijent - Vijesti - Google - Mistranslation - BLEU..... | 47 |
| Slika 71 - Pearsonov koeficijent - Recept - Bing - Accuracy - chrF++ | 48 |
| Slika 72 - Pearsonov koeficijent - Recept - Bing - Mistranslation - chrF++ | 48 |
| Slika 73 - Pearsonov koeficijent - Recept - Google - Accuracy - chrF++..... | 49 |
| Slika 74 - Pearsonov koeficijent - Recept - Google - Mistranslation - chrF++ | 49 |
| Slika 75 - Pearsonov koeficijent - Upute - Bing - Accuracy - chrF++..... | 50 |
| Slika 76 - Pearsonov koeficijent - Upute - Bing - Mistranslation - chrF++ | 50 |
| Slika 77 - Pearsonov koeficijent - Upute - Google - Accuracy - chrF++..... | 51 |
| Slika 78 - Pearsonov koeficijent - Upute - Google - Mistranslation - chrF++ | 51 |
| Slika 79 - Pearsonov koeficijent - Vijesti - Bing - Accuracy - chrF++..... | 52 |
| Slika 80 - Pearsonov koeficijent - Vijesti - Bing - Mistranslation - chrF++ | 52 |
| Slika 81 - Pearsonov koeficijent - Vijesti - Google - Accuracy - chrF++ | 53 |
| Slika 82 - Pearsonov koeficijent - Vijesti - Google - Mistranslation - chrF++ | 53 |

POPIS TABLICA

| | |
|---|----|
| Tablica 1- prikaz rezultata BLEU metrike..... | 10 |
| Tablica 2- prikaz rezultata chrF++ metrike | 14 |
| Tablica 3 - Prikaz rezultata ljudske evaluacije prema MQM metrici..... | 19 |
| Tablica 4 - Prikaz rezultata Pearsonove korelacije – kategorija Accuracy..... | 54 |
| Tablica 5 - Prikaz rezultata Pearsonove korelacije - kategorija Mistranslation..... | 54 |
| Tablica 6 - Pearsonov koeficijent korelacije - rezultati..... | 55 |