

# Strojno prevajanje in slovenščina

Jernej Vičič\*

\*Primorski Inštitut za Naravoslovje in Tehnologijo, Univerza na Primorskem  
6000 Koper  
jernej.vicic@upr.si

## Povzetek

Članek predstavlja pregled strojnih prevajalnih sistemov, ki omogočajo prevajanje v slovenski jezik ali iz slovenskega jezika. Osnovo članka predstavlja obsežna primerjava vseh sistemov, ki podpirajo prevode v ali iz slovenskega jezika. Uporabljene metodologije evalvacije kakovosti prevodov so bile: samodejni objektivni metodi na osnovi metrik BLEU in METEOR ter subjektivni ročni metodi, prva temelječa na metriki WRR (Word-Recognition Rate) ter metoda, temelječa na smernicah LDC. Primerjani sistemi so: Google translate, Microsoft Bing translator, Amebis Presis ter GUAT. Rezultati raziskave so pokazali velik napredek sistemov temelječih na statističnih metodah, vendar ti sistemi kažejo sistematske napake, ki jih bo z opisanimi metodami težko odpraviti.

## Machine Translation and Slovenian Language

The paper presents an overview of the machine translation systems that deal with Slovenian language. It is based on a comprehensive evaluation and comparison of all the publicly available machine translation systems that include Slovenian language in translation pairs. Evaluation methodologies used in the experiment were: automatic evaluation based on metrics BLEU and METEOR and subjective assisted evaluation based on metric WRR (Word-Recognition Rate) and LDC guidelines. The systems included in the comparison were: Google translate, Microsoft Bing translator, Amebis Presis and GUAT.

### 1. Uvod

Članek predstavlja pregled strojnih prevajalnih sistemov, ki omogočajo prevajanje v slovenski jezik ali iz slovenskega jezika. Kakovost strojnega prevajanja je v zadnjih letih zelo napredovala, s prihodom velikih korporacij pa je tudi dostopnost prevajalnih sistemov olajšana. Članek v osnovi poskuša odgovoriti na jasno vprašanje: kateri prevajalni sistem je najboljši, vendar se je med izvedbo eksperimentov pokazalo, da odgovor na to vprašanje ni nedvoumen. Predstavljeni so vsi sistemi, ki med podprtimi jezikovnimi pari ponujajo slovenščino. Pri primerjavi sistemov so bile uporabljene samodejne ter ročne metodologije evalvacije kakovosti prevodov, rezultati kažejo, da je korelacija med samodejnimi metodami ter ročnim testiranjem pri nekaterih sistemih zelo nizka.

Izdelan je bil povedno poravnan korpus besedil, ki je bil uporabljen v testne namene pri vseh metodologijah evalvacije.

Nadaljevanje članka je razdeljeno, kot sledi: v Razdelku 2. so prikazani razlogi za izbiro jezikovnih parov, uporabljenih v eksperimentu, v podrazdelkih je vsak jezikovni par obširneje predstavljen z vidika strojnega prevajanja. Razdelek 3. predstavlja izbiro prevajalnih sistemov ter obširne opise uporabljenih tehnologij pri posameznih sistemih. Sledi Razdelek 4., ki predstavlja metodologijo testiranja ter samo testiranje in rezultate testiranja. V Razdelku 5. so predstavljeni osnovni izsledki, nakazano je tudi možno nadaljnje delo.

### 2. Jezikovni pari

Evalvacijo smo razdelili na dve skupini; na testiranje sistemov s prosto izbiro jezikovnih parov ter na sisteme

sorodnih jezikov. Za jezikovne pare prve skupine smo izbrali skupni jezikovni par slovenščina-angleščina, ki je v povezavi s slovenskim jezikom tudi največkrat uporabljan. Med sorodnimi jezikovnimi pari smo izbrali jezikovni par slovenščina-srbščina, ki edini omogoča primerjavo med sistemi, saj edini slovenski sistem podpira le ta jezikovni par.

#### 2.1. Slovenščina-angleščina

Slovenščina in angleščina nista sorodna jezika, najpomembnejše razlike so na morfološkem ter skladenjskem nivoju. Angleščina je jezik z dokaj omejeno morfologijo ter relativno fiksnim besednim vrstnim redom. Z vidika strojnega prevajanja (Machine Translation - MT) je pomembno, da lahko angleščina kaže večjo stopnjo dvoumnosti posameznih besednih oblik v odvisnosti od njihovih morfoloških oznak, še posebej besednih vrst; ista besedna oblika je lahko samostalnik, glagol ali kaj drugega. To se lahko zgodi tudi v slovenščini, vendar v veliko manjšem obsegu. Druga lastnost angleščine, ki vpliva na kakovost prevodov, je njena sposobnost za gradnjo zapletenih samostalniških zloženkov (npr. long term car park courtesy pickup vehicle), medtem ko slovenščina (in tudi drugi slovanski jeziki) veliko pogosteje uporabljajo kombinacijo pridevnikov in samostalnikov (dostavno vozilo dolgoročnega parkirišča), kar pomeni, da je običajno potrebno prevesti pridevnik kot samostalnik ali obratno, kar je seveda izziv za vsak strojni prevajalni sistem.

#### 2.2. Slovenščina-srbščina

Tako slovenščina kot srbščina pripadata skupini južno-slovanskih jezikov, ki jih govorijo večinoma prebivalci z območja bivše Jugoslavije. Srbščina je najbolj razširjena v Srbiji, slovenščina v Sloveniji. Jezika si delita skupne ko-

renine ter še pomembneje skupno polpreteklo zgodovino; oba jezika sta bila uradna jezika skupne države, celo predavana v šolah kot materina jezika ter jezika okolja.

Obstaja niz pomembnih razlogov za postavitev prevajalnega sistema za predstavljen jezikovni par, omenimo le najpomembnejša: ekonomiji obeh novih držav, Slovenije ter Srbije, sta še vedno tesno povezani; mlajše generacije imajo jezikovne težave v medsebojni komunikaciji.

Oba jezika sta visoko pregibna ter morfološko bogata.

### 2.3. Izbira in priprava testnega gradiva

Začetni načrt testiranja je predvideval uporabo naključno izbranih povedi vzporedno poravnane večjezičnega dela korpusa MULTEXT-EAST (Dimitrova et al., 1998), romana 1984 Georga Orwella. Ta korpus vsebuje poravnane povedi v vseh treh jezikih, ki smo jih nameravali uporabiti pri testiranju.

Prvi krog testiranja je pokazal, da je Google zelo verjetno uporabil korpus (Dimitrova et al., 1998) pri učenju svojega sistema. Podobnost prevodov sistema z referenčnimi prevodi je bila več kot le naključna.

Odločili smo se, da bomo zamenjali testno gradivo. Uporabili bi lahko druge večjezične korpuske kot je na primer Opus (Tiedemann, 2009). Izključili smo vse korpuske, ki so dostopni prek interneta, saj pri teh korpusih obstaja možnost, da so bili uporabljeni kot učno gradivo pri izdelavi prevajalnih sistemov.

Pripravili smo novo učno množico, ki je bila zgrajena iz podnapisov filma, izbrali smo Matrix. Ta film je bil, in je še, zelo priljubljen med piratskimi uporabniki, tako smo lahko pričakovali prevode v vseh zelenih jezikih (slovenščina, angleščina ter srbsščina) in tudi dovolj kakovostne prevode, saj skupnost večih uporabnikov omogoča preverjanje ter popravljanje napak v prevodih ter ocenjevanje posameznih prevodov. Izbrali smo različice prevodov z najboljšimi ocenami ter največ prenosi ter jih še ročno pregledali.

Podnapisi filmov že vsebujejo informacijo o času prikaza na zaslonu, kar omogoča enostavno izdelavo poravnanih korpusov. Izdelali smo povedno poravnani korpus dveh jezikovnih parov z enim skupnim jezikom. Pri poravnavi smo uporabili večino metod, opisanih v (Tiedemann, 2007). Segmente, kjer je bilo časovno ujemanje podnapisov dvoumno, smo izpustili. Podnapisi filmov imajo v povprečju krajše, posledično strukturno manj zapletene povedi. To dejstvo smo omejili z izločitvijo najkrajših povedi, mejo smo postavili empirično pri 20 znakih. Tako smo dobili korpus z osnovnimi lastnostmi, predstavljenimi v Tabeli 1

Tabela 1: Osnovni podatki o testnem korpusu

jezikovni par	jezik	št. povedi	št. besed
sl-en	slovenščina	760	4814
sl-en	angleščina	760	5646
sl-sr	slovenščina	742	4624
sl-sr	srbsščina	742	4460

## 3. Pregled sistemov

Po pregledu spleta ter pogovoru s strokovnjaki področja smo se odločili za naslednji izbor prevajalnih sistemov:

Google Translate, Microsoft BING translator, Amebis Presis ter GUAT. Vrstni red sistemov je naključen. Vse izbrane sisteme lahko uvrstimo v dve paradigmi strojnega prevajanja.

Google Translate ter Microsoft BING translator spadata v paradigmo sistemov statističnega strojnega prevajanja (Statistical Machine Translation - SMT). Takšni sistemi so osnovani na parametričnih statističnih modelih, ki so naučeni na poravnanih dvojezičnih korpusih (učnih primerih). Namesto razdeljevanja stavkov po slovničnih pravilih iščemo splošne vzorce, ki se porajajo pri uporabi jezika. Glavna prednost statističnega pristopa je, da so metode neodvisne od jezika (čeprav uporabnejše za določene jezike, med temi ni slovenščine). Glavna pomanjkljivost sistemov strojnega prevajanja na osnovi pravil je slab pregled nad delovanjem sistema, sistematske napake je zelo težko odpraviti, uvajanje lingvističnega znanja je le delno mogoče oziroma celo nemogoče.

Amebis Presis ter GUAT spadata med sisteme strojnega prevajanja temelječe na pravilih (Rule-Based Machine Translation - RBMT). Način zapisa pravil se razlikuje med sistemi, veže pa jih skupno dejstvo, da je postavitev takšnega sistema dolgotrajno opravilo. V to skupino sodi večina današnjih komercialnih prevajalnih sistemov, čeprav se pri gradnji poslužujejo nekaterih manj standardnih prijemov. Sistemi te paradigme izvorno besedilo najprej morfološko ter skladenjsko analizirajo ter izdelajo predstavitev vhodnega besedila, ponavadi v obliki skladenjskega drevesa izpeljave. Ta predstavitev se še dodatno abstrahira s poudarkom na zahtevah strojnega prevajanja. Proces transferja prevede abstraktno predstavitev vhodnega besedila v izvornem jeziku v podobno predstavitev v ciljnem jeziku, to predstavitev sistem uporabi kot osnovo za generacijo besedila v ciljnem jeziku, v bistvu uporabi inverzne metode prvega dela na ciljnem jeziku.

### 3.1. Google Translate

Google Translate<sup>1</sup> je tipični pripadnik sistemov statističnega strojnega prevajanja (Statistical Machine Translation - SMT), ki je predstavljena v Razdelku 3.. Sistem ne uporablja dodatnega jezikovnega znanja, zanaša se samo na korelacijo med znanimi pari že prevedenih vzporednih besedil. Statistične metode zahtevajo ogromne količine besedil ter veliko računalniške moči za obdelavo teh besedil. Google ima oboje, besedila, nabrana za izdelavo iskalnika, ter veliki grozdi računalnikov omogočajo hitro izdelavo sistemov za strojno prevajanje z zavirljivo kakovostjo.

Opisane lastnosti so omogočile Googlu izdelavo prevajalnih sistemov za kar 58 svetovnih jezikov<sup>2</sup>, torej za kar  $58 * 58 = 3364$  jezikovnih parov.

O natančnem delovanju prevajalnega sistema Google Translate ni veliko znanega, znane so le osnovne metode. Pri uporabi sistema za prevode ter pri izvajanju evalvacije prevajalnega sistema smo prišli do podobnih zaključkov kot nekaj avtorjev, ki je svoja mnenja predstavilo na odprti debati na dopisnem seznamu Corpora (List, 2010).

<sup>1</sup>Google Translate: <http://translate.google.com>

<sup>2</sup><http://translate.google.com>, podatek z dne: 20.6.2010

Za večje jezike ter pogostejše jezikovne pare, kot so angleščina-nemščina, angleščina-francoščina ali angleščina-kitajščina, so sistemi naučeni posebej, medtem ko so manjši jeziki, mednje sodi tudi slovenščina, prevajani prek vmesnega jezika. To pomeni, da prevajanje v slovenščino ne poteka neposredno iz izvornega jezika, ampak se najprej besedilo prevede v angleščino (ponovno špekulacija) ter šele nato v slovenščino. Kakovost prevodov je slabša kot pri večjih jezikovnih parih, kar potrjujejo tudi rezultati testiranja, prikazani v Razdelku 4. Predstavljeno špekulacijo lahko potrdimo le z empiričnimi primeri prevodov, kot je primer prikazan na Sliki 1

Razumeš?

Гот ит?

Romanizirano: Got it?

Slika 1: Primer prevoda iz slovenščine v srbsščino, kjer se v prevodu pojavljajo angleške besede. Primer kaže na verjetno uporabo angleščine kot vmesnega jezika pri prevajanju.

### 3.2. Microsoft BING Translator

Bing Translator<sup>3</sup> (Quirk et al., 2005) je hibridni sistem za strojno prevajanje naravnih jezikov. Sistem temelji na statističnem strojnem prevajalniku, obširneje je predstavljen v Razdelku 3., ki uporablja tudi pravila, ki so odvisna od jezika ter določeno mero analize izvornega besedila. Microsoft imenuje ta sistem kot "jezikovno obveščeno statistično strojno prevajanje" (Linguistically informed statistical machine translation).

Sistem je v osnovi statistični sistem za strojno prevajanje na osnovi fraz, ki vključuje jezikovno odvisno analizo besedila, drevesa odvisnosti (dependency trees) ter drevesa izpeljave (parse trees) in pravila za poravnavo besed (word alignment rules) za generalizacijo naučenih fraz.

### 3.3. Amebis Presis

Prevajalni sistem Presis podjetja Amebis (Romih in Hožan, 2002) je bil prvi sistem za strojno prevajanje, ki je med prevajalnimi jezikovnimi pari vseboval slovenski jezik.

Sistem sodi v paradigmo strojnih prevajalnih sistemov na osnovi pravil (Rule-Based Machine Translation - RBMT), natančneje je predstavljena v Razdelku 3., Presis analizira vsako poved v izvornem jeziku v slovnične komponente, kot so osebek, predmet, povedek in atributi ustreznih semantičnih kategorij. Na osnovi analiziranega izvornega besedila izbere pripravljena pravila, ki omogočajo prevod analiziranih komponent v ciljni jezik, nato sintetizira poved v ciljnem jeziku.

Ena glavna odlika sistema Presis je možnost prilagoditve prevajalni domeni z vključitvijo dodatnih slovarjev vendar primerjava tako spremenjenega sistema ne bi bila objektivna, zato smo to možnost opustili.

<sup>3</sup>Microsoft Bing Translator: <http://www.microsofttranslator.com/>

Pri testiranju smo uporabili spletno različico prevajalnega sistema Amebis Presis 2.0-pre37<sup>4</sup>

### 3.4. GUAT

GUAT<sup>5</sup> je rezultat akademskih eksperimentov metod za hitro postavitev prevajalnih sistemov za sorodne jezike. Izdelava tega sistema je opisana v (Vičič, 2009) in (Vičič in Homola, 2010). Osnova sistema je odprtokodno ogrodje Apertium (Corbi-Bellot et al., 2005). Apertium je ogrodje za izdelavo za sistemov za strojno prevajanje na osnovi pravil plitkega prenosa (Shallow Transfer Rule-Based Machine Translation - RBMT). Okolje je posebej primerno za sorodne jezike, saj omogoča le morfološko analizo ter sintezo besedil, ne vsebuje pa orodij za popolno analizo izvornih besedil. Prevajalni sistemi za skladiščenje različne jezike, kot sta slovenščina in angleščina bi dosegali slabše rezultate.

### 3.5. Ostala orodja

Pri pregledu obstoječih sistemov smo naleteli še na dva prevajalna sistema, ki vsebujeta slovenščino med prevajalnimi pari: VoiceTran<sup>6</sup> (Žganec Gros et al., 2005) ter Prevajalnik.net<sup>7</sup>.

Prvi je glasovni komunikator in je najbolj uporaben za podajanje povelj ter postavljanje preprostih vprašanj. Omogoča prepoznavanje govora ter ročne pisave. Del sistema VoiceTran je tudi sistem za prevajanje besedila. Zaradi ožje namembnosti sistema bi bila smiselnost takšnega testiranja vprašljiva. Sistem temelji na posebni različici orodja Presis, ki je opisano v Razdelku 3.3.

Prevajalnik.net je spletna storitev, ki omogoča prevode iz slovenščine ter v slovenščino za več kot 40 svetovnih jezikov. Storitev uporablja Googlovem prevajalnik, ki je opisan v Razdelku 3.1. Za testiranje tega sistema se nismo odločili, saj bi bili rezultati identični Googlovim.

## 4. Metodologija vrednotenja ter rezultati

Testni primeri, uporabljeni pri testiranju so predstavljeni v Razdelku 2.3., za ponovljivost rezultatov raziskave so vsa gradiva na voljo na naslovu: [http://jt.upr.si/mt\\_v\\_sloveniji/](http://jt.upr.si/mt_v_sloveniji/). Pri vrednotenju so bile uporabljene štiri metode:

1. Samodejno vrednotenje s pomočjo metrike BLEU (Papinen et al., 2001).
2. Samodejno vrednotenje s pomočjo metrike METEOR (Lavie in Agarwal, 2007).
3. Ročno vrednotenje, temelječe na metriki WRR (Word-Recognition Rate)
4. Ročno vrednotenje, temelječe na smernicah LDC (LDC, 2005).

Metriki BLEU ter METEOR sta najbolj razširjeni samodejni metriki za evalvacijo strojnega prevajanja, odločili

<sup>4</sup>Presis, <http://presis.amebis.si/>

<sup>5</sup>Prevajalni sistem GUAT: <http://jt.upr.si/guat/>

<sup>6</sup>VoiceTRAN: <http://www.voicetran.org/>

<sup>7</sup>VoiceTRAN: <http://prevajalnik.net/>

smo se za uporabo obeh, čeprav se mnogi avtorji strinjajo, da metrika BLEU ni primerna za primerjavo različnih sistemov (Callison-Burch et al., 2006). Pri uporabi samodejnih metrik smo izvirne povedi poravnane korpusa predstavljene v Razdelku 1 prevedli s prevajalnim sistemom. Prevode sistemov in referenčne prevode, ciljne povedi testnega korpusa, smo primerjali z obema metrikama. Uporabili smo vse razpoložljive testne primere, osnovni podatki o testnem korpusu so zapisani v Razdelku 1.

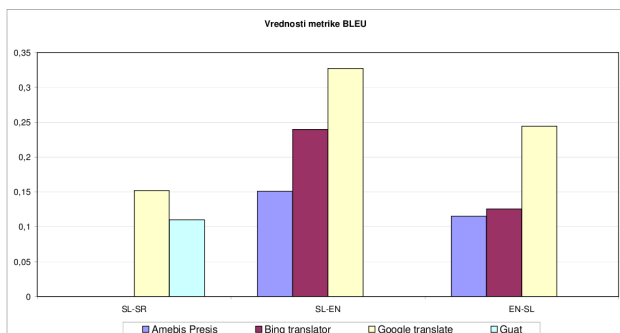
Obe metodi ročnega vrednotenja sta bili izvedeni na podoben način: evalvacija je obsegala naključno izbiro 100 povedi iz obsega vseh testnih povedi, prevod teh povedi s prevajalnim sistemom in izvedbo ročne evalvacije. Testiranje smo opravili s po dvema ocenjevalcema za vsak ciljni jezik. Slovenščina je bila obema ocenjevalcema materin jezik, za angleščino pa smo uporabili enega ocenjevalca, ki mu je angleščina materin jezik in enega, ki aktivno obvlada ta jezik.

#### 4.0.1. Samodejno vrednotenje s pomočjo metrike BLEU

Bilingual Evaluation Understudy - BLEU (Papineni et al., 2001) je bila prva in še vedno najbolj razširjena metrika za evalvacijo kakovosti prevodov sistemov strojnega prevajanja. Kakovost prevodov je predstavljena kot natančnost ujemanja prevodov sistema za strojno prevajanje z referenčnimi prevodi poklicnih prevajalcev. Vrednosti so izračunane za posamezne prevedene odseke, ponavadi povedi, ter povprečene za celoten testni korpus. Berljivost ter slovnična pravilnost nista upoštevani.

BLEU uporablja spremenjeno različico preciznost (precision), ki za razred predstavlja število pravilno klasificiranih elementov (true positives), za primerjavo kandidata za prevod z enim ali več referenčnimi prevodi. Sprememba z osnovno preciznost naj bi poskrbela za lastnost sistemov strojnega prevajanja, ki težijo k daljšim prevodom.

Uporabili smo javno dostopno implementacijo metrike BLEU (NIST, 2008), različico v11b. Rezultati so prikazani na Sliki 2, večje vrednosti predstavljajo boljše rezultate. Rezultati testiranja z metriko BLEU kažejo na ve-



Slika 2: Rezultati testiranja z metriko BLEU. Sistem Google dosega najboljše rezultate.

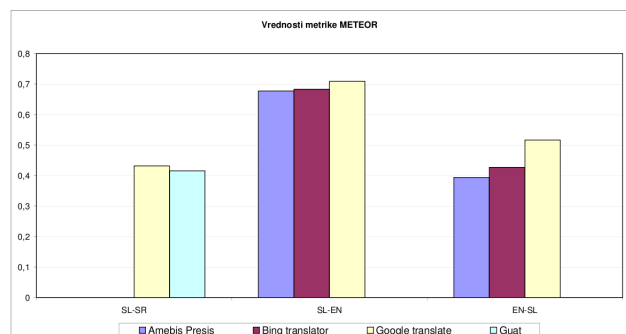
liko prednost sistema Google translate. Pomisleki avtorjev

glede predstavljenih rezultatov metrike BLEU so predstavljeni v Razdelku 5.

#### 4.0.2. Samodejno vrednotenje s pomočjo metrike METEOR

Uporabili smo javno dostopno implementacijo metrike METEOR (Lavie in Agarwal, 2007), različico 0.6. Metrika temelji na harmonični sredini natančnosti ter priklica unigramov (unigram precision and recall), kjer je priklic močnejše utežen kot natančnost. Vsebuje še več metod jezikovnih tehnologij, ki niso prisotne pri ostalih samodejnih metrikah strojnega prevajanja kot so krnjenje in ujemanje sinonimov kot pomoč pri iskanju ujemanja besed. Krnjenje je predvsem primerno za visiko pregibne jezike saj omejuje vpliv napačne uporabe pregibanja; na primer napačne uporabe sklona pri samostalnikih.

Samo orodje nima priložene kode za krnjenje srbskega jezika, uporabili smo orodje, ki je bilo izdelano za vrednotenje rezultatov metod, opisanih v (Vičič in Homola, 2010). Rezultati so predstavljeni na Sliki 3, večje vrednosti predstavljajo boljše rezultate.

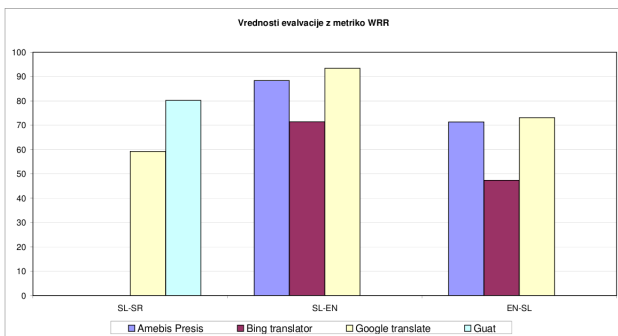


Slika 3: Rezultati metrike METEOR.

#### 4.0.3. Ročno vrednotenje, temelječe na metriki WRR (Word-Recognition Rate)

Metrika temelječa na utežena Levenshteinovi razdalji (weighted Levenshtein edit-distance) (Levenshtein, 1965), poznana tudi kot Word Error Rate (WER) izračuna najmanjše število sprememb, ki jih moramo narediti za izdelavo *pravilne* povedi v ciljnem jeziku iz samodejno izdelane povedi (prevoda ocenjevanega sistema). Število sprememb še utežimo z dolžino povedi. Dovoljene spremembe so vstavitev, brisanje ter zamenjava besede.

Sama izvedba evalvacije je bila sestavljena iz samodejnega prevoda testnih primerov ter preštevanje števila sprememb, ki jih moramo opraviti za izdelavo dovolj dobrega prevoda. Definicija dovolj dobrega prevoda, ki smo jo uporabili pri tem eksperimentu, je prevod, ki je sintaktično pravilen ter izraža poln pomen izvirne povedi. Rezultati na Sliki 4 predstavljajo WRR, Word Recognition rate (1 - WER), ki predstavlja uspešnost prevajalnega sistema in ne velikosti napake sistema, torej večje vrednosti so boljše.



Slika 4: Rezultati evalvacije z metriko Word Recognition Rate - WRR. Pri jezikovnem paru s srbščino dosega GUAT signifikantno boljše rezultate kot Google, pri jezikovnih parih z angleščino so signifikantno slabši rezultati sistem Bing, razlika med ostalima sistemoma ni signifikantna.

#### 4.0.4. Ročno vrednotenje, temelječe na smernicah LDC

Subjektivno ročno ocenjevanje kakovosti prevajalnih sistemov smo izvedli po smernicah določenih na delavnici NIST Machine Translation Evaluation Workshop konzorcija Linguistic Data Consortium - LDC (LDC, 2005). Ta metodologija je najbolj razširjena pri ročni evalvaciji sistemov strojnega prevajanja.

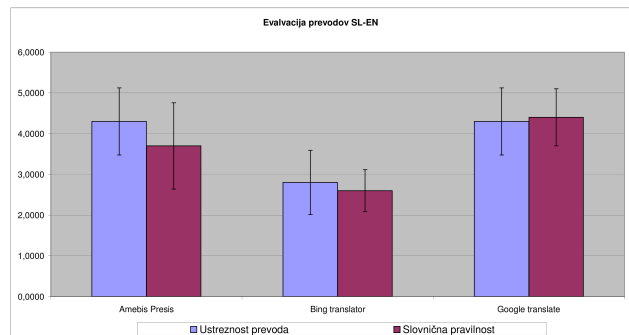
Uporabljeni sta bili dve pet-točkovni lestvici, ena opisuje ustreznost prevodov (adequacy), druga pa slovnično pravilnost prevodov (fluency). Lestvica ustreznosti prevodov (adequacy) kaže, koliko pomena iz referenčne povedi je izraženega v hipotetičnem prevodu: 5 = Vse/All, 4 = Večina/Most, 3 = Veliko/Much, 2 = Malo/Little, 1 = Nič/None.

Druga lestvica kaže slovnično pravilnost prevodov in je vsebina ne zanima. Vrednosti lestvice ustrezajo: 5 = Slovnično popoln prevod/Flawless translation, 4 = Slovnično dobro besedilo/Good target language, 3 = Besedilo ni materin jezik/Non-native target language, 2 = Slovnično nepravilno besedilo/Disfluent target language, 1 = Nerazumljivo besedilo/Incomprehensible text.

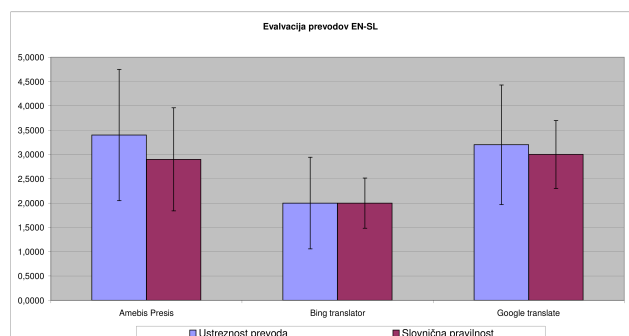
Ločeni lestvici za ustreznost prevodov ter slovnično pravilnost sta bili razviti pod predpostavko, da je lahko prevod slovnično nepravilen, pa še vedno vsebuje ves pomen izvornega besedila in v določenih primerih to že zadošča.

Rezultati so predstavljeni na Slikah 5, 6, 7 Vsaka slika kaže primerjavo sistemov za določen jezikovni par. Standardna deviacija kaže odstopanja vrednotenja posameznih povedi. Večje vrednosti predstavljajo boljše rezultate.

Tabela 2 kaže zadostno mero ujemanja do zelo veliko mero ujemanja med ocenjevalci (satisfactory to very high inter-rater agreement) po Cohenovem koeficientu kappa (Cohen, 1960).



Slika 5: Rezultati evalvacije sistemov jezikovnega para SL-EN po smernicah (LDC, 2005). Sistem Presis dosega rezultate primerljive z Googlom, sistem Bing dosega signifikantno slabše rezultate.

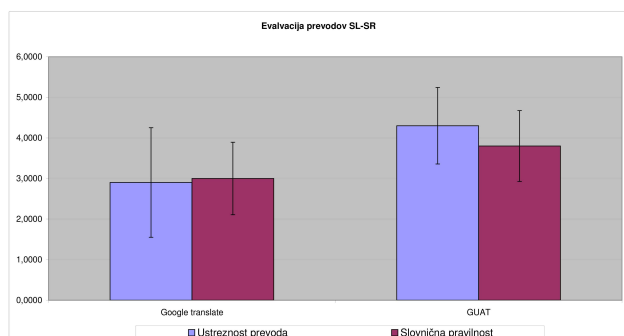


Slika 6: Rezultati evalvacije sistemov jezikovnega para EN-SL po smernicah (LDC, 2005). Sistem Presis dosega rezultate primerljive z Googlom, sistem Bing dosega signifikantno slabše rezultate.

## 5. Zaključek in nadaljnje delo

Začetni načrt evalvacije je obsegal vse opisane metode evalvacije, čeprav smo se zavedali neprimernosti metrike BLEU (Papineni et al., 2001) za primerjavo različnih prevajalnih sistemov. Veliko avtorjev se strinja, da metrika BLEU sistematično zapostavlja sisteme RBMT, kot na primer (Callison-Burch et al., 2006; Labaka et al., 2007), prav tako metrika ni primerna za visoko pregibne jezike. Dva od predstavljenih sistemov spadata v skupino prevajalnih sistemov paradigme RBMT in sta se pri tej metriki zelo slabo odrezala. Evalvacijo z BLEU metriko smo vseeno uvrstili v eksperiment zaradi zgodovinske primerljivosti rezultatov. Druga uporabljena samodejna metrika METEOR je po priporočilu avtorjev veliko primernejša za visoko pregibne jezike in odpravlja največje pomanjkljivosti metrike BLEU, rezultati še vedno postavljajo Googlov prevajalni sistem na prvo mesto, vendar z manjšo razliko.

Ročni metodologiji, ki veliko bolje odražata dejansko kakovost prevodov, kažeta veliko boljše rezultate sistemov RBMT, vendar moramo upoštevati manjše število testnih primerov ter na možno neobjektivnost evalvatorjev, čeprav je bilo ujemanje med evalvatorji zadovoljivo do odlično. Primerjava sistemov jezikovnih kombinacij SL-EN ter EN-SL kaže na veliko slabšo kakovost prevodov v iz angleščine



Slika 7: Rezultati evaluacije sistemov jezikovnega para SL-SR po smernicah (LDC, 2005). Sistem GUAT dosega signifikantno boljše rezultate kot Google.

Tabela 2: Cohenov koeficient kappa (Cohen, 1960) za posamezne sisteme kaže zadostno mero ujemanja do zelo veliko mero ujemanja med ocenjevalci. Prikazani so samo podatki primerjave ujemanja ocenjevalcev sistemov jezikovnega para sl-en

	Presis	Bing	Google
kappa	0,86	0,69	0,78
pričak. naklj. ujemanje	0,300	0,317	0,305
št. primerov	100	100	100

v slovenščino v primerjavi s prevodi iz slovenščine v angleščino. Podajamo možno razlago, ki pa pojasni le del problema: statistični prevajalni sistemi za izbiro končnih kandidatov uporabljajo jezikovne modele ciljnih jezikov, jezikovni model za slovenščino je naučen na veliko manjši učni množici kot jezikovni model za angleščino, učno množico pogojuje število dostopnih besedil v elektronski obliki. Ta razlaga je brezpredmetna pri sistemu Presis, ki temelji na pravilih.

Pregled osnovnih napak sistemov: glavna napaka statističnih sistemov je problem lokalnega ujemanja besed v leksikalnih kategorijah. Kot primer navedimo ujemanje samostalnikov ter pridevnikov v spolu, sklonu ter številu. Ta problem je še posebej izrazen v slovanskih jezikih. Pri sistemih paradigme RBMT je opazno manjše besedišče. Ta problem je še posebej prisoten pri sistemu GUAT, ki ima dvojezični slovar ter slovar ciljnega jezika (srbsčine) izluščen iz relativno majhnega korpusa. Sistemi za strojno prevajanje se nenehno izboljšujejo, primerjava prikazana v tem članku bo ob rednih podobitvah dostopna na naslovu: [http://jt.upr.si/mt\\_v\\_sloveniji/](http://jt.upr.si/mt_v_sloveniji/).

## 6. Literatura

- Chris Callison-Burch, Miles Osborne, in Philipp Koehn. 2006. Re-evaluating the role of Bleu in machine translation research. V: *Proceedings of EACL*.
- Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.
- Antonio M. Corbi-Bellot, Mikel L. Forcada, Sergio Ortiz-

- Rojas, Juan Antonio Prez-Ortiz, Gemma Ramirez-Sanchez, Felipe Sanchez-Martinez, Inaki Alegria, Aingeru Mayor, in Kepa Sarasola. 2005. An open-source shallow-transfer machine translation engine for the Romance languages of Spain. V: *Proceedings EAMT conference*, str. 79–86, May.
- Ludmila Dimitrova, Nancy Ide, Vladimir Petkevič, Tomaž Erjavec, Heiki Jaan Kaalep, in Dan Tufis. 1998. Multext-East: Parallel and Comparable Corpora and Lexicons for Six Central and Eastern European Languages. V: *COLING-ACL*, str. 315–319.
- Gorka Labaka, Nicholas Stroppa, Andy Way, in Kepa Sarasola. 2007. Comparing rule-based and data-driven approaches to Spanish-to-Basque machine translation. V: *Proceedings of the Machine Translation Summit XI*, str. 41–48.
- A. Lavie in A. Agarwal. 2007. METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments. V: *Proceedings of Workshop on SMT at the ACL conference*.
- LDC. 2005. Linguistic data annotation specification: Assessment of fluency and adequacy in translations. Tehnično poročilo, LDC.
- V. Levenshtein. 1965. Binary codes capable of correcting deletions, insertions and reversals. *Doklady Akademii Nauk*, str. 845–848.
- Corpora List. 2010. Corpora list.
- NIST. 2008. Evaluation software.
- Kishore Papineni, Salim Roukos, Todd Ward, in Wei-Jing Zhu. 2001. Bleu: a method for automatic evaluation of machine translation. Tehnično poročilo, IBM.
- Chris Quirk, Arul Menezes, in Colin Cherry. 2005. Dependency treelet translation: Syntactically informed phrasal smt. V: *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*.
- Miro Romih in Peter Holožan. 2002. A slovenian-english translation system. V: *Proceedings of the 3rd Language Technologies Conference*, str. 167.
- Jorg Tiedemann. 2007. Improved sentence alignment for movie subtitles. V: *Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP'07)*, Borovets, Bulgaria.
- Jörg Tiedemann. 2009. News from opus - a collection of multilingual parallel corpora with tools and interfaces. *Recent Advances in Natural Language Processing*, 5:237–248.
- Jernej Vičič in Petr Homola. 2010. Speeding up the implementation process of a shallow transfer machine translation system. V: *Proceedings of the 14th EAMT Conference*, str. 261–268, Saint Raphael, France. European Association for Machine Translation.
- Jernej Vičič, 2009. *Metode hitre izdelave gradiv za prevajalne sisteme plitkega prenosa za visoko pregibne jezike*, str. 133–153. Znanstveno-raziskovalno središče, Založba Annales.
- Jerneja Žganec Gros, France Mihelič, Tomaž Erjavec, in Špela Vintar, 2005. *The VoiceTRAN Speech-to-Speech Communicator*, str. 379–384. Springer Berlin / Heidelberg.