

Metode vrednovanja strojnih prijevoda prema kriteriju razumljivosti

Ljubas, Sandra

Doctoral thesis / Doktorski rad

2023

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

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:162:353957>

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

Download date / Datum preuzimanja: **2024-07-21**



Sveučilište u Zadru
Universitas Studiorum
Jadertina | 1396 | 2002 |

Repository / Repozitorij:

[University of Zadar Institutional Repository](#)

SVEUČILIŠTE U ZADRU

POSLIJEDIPLOMSKI SVEUČILIŠNI STUDIJ
HUMANISTIČKE ZNANOSTI

Sandra Ljubas

**METODE VREDNOVANJA STROJNIH
PRIJEVODA PREMA KRITERIJU
RAZUMLJIVOSTI**

Doktorski rad

Zadar, 2023.

SVEUČILIŠTE U ZADRU
POSLIJEDIPLOMSKI SVEUČILIŠNI STUDIJ
HUMANISTIČKE ZNANOSTI

Sandra Ljubas

**METODE VREDNOVANJA STROJNIH
PRIJEVODA PREMA KRITERIJU
RAZUMLJIVOSTI**

Doktorski rad

Mentor:

doc. prof. dr. sc. Božo Bekavac

Komentorica:

prof. dr. sc. Marijana Kresić Vukosav

Zadar, 2023.

SVEUČILIŠTE U ZADRU

TEMELJNA DOKUMENTACIJSKA KARTICA

I. Autorica i studij

Ime i prezime: Sandra Ljubas

Naziv studijskog programa: Poslijediplomski doktorski studij Humanističke znanosti

Mentor: doc. prof. dr. sc. Božo Bekavac

Komentorica: prof. dr. sc. Marijana Kresić Vukosav

Datum obrane: 27. studenoga 2023.

Znanstveno područje i polje u kojem je postignut doktorat znanosti: humanističke znanosti, filologija – jezikoslovlje

II. Doktorski rad

Naslov: Metode vrednovanja strojnih prijevoda prema kriteriju razumljivosti

UDK oznaka: 81'322.4:001.8

Broj stranica: 206

Broj slika/grafičkih prikaza/tablica: 4/12/2

Broj bilježaka: 18

Broj korištenih bibliografskih jedinica i izvora: 109

Broj priloga: 5

Jezik rada: hrvatski

III. Stručna povjerenstva

Stručno povjerenstvo za ocjenu doktorskog rada:

1. prof. dr. sc. Nataša Pavlović, predsjednica
2. izv. prof. dr. sc. Goranka Antunović, članica
3. izv. prof. dr. sc. Vanda Mikšić, članica

Stručno povjerenstvo za obranu doktorskog rada:

1. prof. dr. sc. Nataša Pavlović, predsjednica
2. izv. prof. dr. sc. Goranka Antunović, članica
3. izv. prof. dr. sc. Vanda Mikšić, članica

UNIVERSITY OF ZADAR
BASIC DOCUMENTATION CARD

I. Author and study

Name and surname: Sandra Ljubas

Name of the study programme: Postgraduate doctoral study in Humanities

Mentor: Assistant Professor Božo Bekavac, PhD

Co-mentor: Professor Marijana Kresić Vukosav, PhD

Date of the defence: 27

Scientific area and field in which the PhD is obtained: Humanities, Philology – linguistics

II. Doctoral dissertation

Title: Methods for Evaluating Intelligibility in Machine Translation

UDC mark: 81'322.4:001.8

Number of pages: 206

Number of pictures/graphical representations/tables: 4/12/2

Number of notes: 18

Number of used bibliographic units and sources: 109

Number of appendices: 5

Language of the doctoral dissertation: Croatian

III. Expert committees

Expert committee for the evaluation of the doctoral dissertation:

1. Professor Nataša Pavlović, PhD, chair
2. Associate Professor Goranka Antunović, PhD, member
3. Associate Professor Vanda Mikšić, PhD, member

Expert committee for the defence of the doctoral dissertation:

1. Professor Nataša Pavlović, PhD, chair
2. Associate Professor Goranka Antunović, PhD, member
3. Associate Professor Vanda Mikšić, PhD, member



Izjava o akademskoj čestitosti

Ja, **Sandra Ljubas**, ovime izjavljujem da je moj **doktorski** rad pod naslovom **Metode vrednovanja strojnih prijevoda prema kriteriju razumljivosti** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 15. veljače 2023.

Sadržaj

1. Uvod	8
2. Teorijski okvir	12
2.1 KRATKA POVIJEST STROJNOG PREVOĐENJA	12
2.1.1 Pristupi strojnom prevođenju kroz povijest	14
2.1.2 Neuronsko strojno prevođenje	15
2.1.3 Mitovi, strahovi i predrasude o strojnom prevođenju.....	18
2.2 VREDNOVANJE STROJNIH PRIJEVODA	23
2.2.1 Ljudske metode vrednovanja vs. automatske metrike	24
2.2.2 Kriteriji ljudskog vrednovanja strojnih prijevoda	26
2.3 METODE VREDNOVANJA STROJNIH PRIJEVODA	28
2.3.1 Metoda subjektivne procjene	28
2.3.1.1 <i>Problem niskih razina slaganja kod subjektivnog ocjenjivanja prijevoda</i>	30
2.3.1.2 <i>Kriteriji kvalitete</i>	32
2.3.1.3 <i>Bodovne ljestvice za ocjenjivanje prijevoda</i>	34
2.3.2 Metode provjere razumljivosti testovima razumijevanja	35
2.3.2.1 <i>Definiranje i ispitivanje pojma „razumljivosti“</i>	36
2.3.2.2 <i>Tehnike ispitivanja razumljivosti u istraživanjima kvalitete strojnih prijevoda</i>	38
2.3.2.3 <i>Priprema testova razumijevanja</i>	42
2.3.3 Metoda analize pogrešaka	43
2.3.3.1 <i>Izazovi provedbe analize pogrešaka</i>	43
2.3.3.2 <i>Dobra praksa analize pogrešaka</i>	45
2.3.3.3 <i>Kategorije pogrešaka</i>	47
2.3.3.4 <i>Odnos zabilježenih pogrešaka i percepcije kvalitete</i>	50
2.3.3.5 <i>Metoda redakture strojnog prijevoda</i>	52
3. Cilj i istraživačka pitanja	57
4. Metodologija	60
4.1 MATERIJAL.....	60
4.2 ISPITANICI.....	71
4.3 METODA	73
5. Rezultati i diskusija	84

5.1	METODA SUBJEKTIVNE PROCJENE: REZULTATI I DISKUSIJA (IP1, IP2)	84
5.1.1	Rezultati vrednovanja subjektivnom procjenom (IP1, rezultati).....	84
5.1.2	Raščlanjeni segmenti ili vrednovanje cjeline? (IP1, diskusija)	88
5.1.3	Suodnos ocjena i prijašnjih stavova ispitanika (IP2, rezultati).....	91
5.1.4	Analiza stavova ispitanika (IP2, diskusija)	93
5.2	METODA TESTA RAZUMIJEVANJA: REZULTATI I DISKUSIJA (IP3)	96
5.2.1	Broj točnih odgovora na testu razumijevanja	97
5.2.2	Suodnos broja točnih odgovora s kvalitetom strojnog prijevoda	99
5.2.3	Proporcionalnost broja točnih odgovora na testu razumljivosti s ocjenama razumljivosti	106
5.3	METODA ANALIZE POGREŠAKA: REZULTATI I DISKUSIJA (IP4, IP5).....	108
5.3.1	Kategorije anotiranih pogrešaka (rezultati i analiza primjera)	108
5.3.2	Utjecaj broja jezičnih pogrešaka na ocjene razumljivosti (IP4, rezultati i diskusija)	116
5.3.3	Utjecaj vrste jezičnih pogrešaka na ocjene razumljivosti (IP5, rezultati i diskusija)	119
5.4	DIJAKRONIJSKA USPOREDBA REZULTATA PO POJEDINIM METODAMA VREDNOVANJA (IP6, REZULTATI I ANALIZA)	124
5.5	USPOREDBA ISTRAŽENIH LJUDSKIH METODA VREDNOVANJA (ZAVRŠNA DISKUSIJA)..	135
6.	Zaključak	140
	Literatura	146
	Sažetak	162
	Abstract	164
	Prilog A	166
	Prilog B	166
	Prilog C	183
	Prilog D	192
	Prilog E	200
	Životopis	206

1. Uvod

Povijest strojnog prevođenja, od memoranduma iz 1949. godine u kojemu je Warren Weaver prevođenju pristupio kao kriptografskom problemu, pa sve do najnovijeg pristupa koji se temelji na neuronskim mrežama i načelima dubokog učenja, u stopu prati i povijest vrednovanja strojnih prijevoda. I konvencionalni se prijevodi općenito svakodnevno podvrgavaju različitim oblicima vrednovanja: prije nego što prevoditelji pošalju svoje prijevode, sami procjenjuju zadovoljavaju li sve zahtjeve kvalitete; sveučilišni profesori ocjenjuju prijevode studenata koji se tek obučavaju za prevoditeljsko zanimanje, upozoravajući ih na pogreške i zamke, te hvaleći vrline prijevoda; poslodavci vrednuju probne prijevode kandidata koji se prijavljuju za prevoditeljske poslove, a donose i odluku je li kvaliteta pojedinih prijevoda dostatna za širu uporabu ili objavljivanje, u kojem slučaju i javnost ima priliku donijeti svoj sud o toj kvaliteti (Yusof, Darus i Aziz 2017: 251).

Javnost vrednovanju pristupa ne uzimajući u obzir razlike između takozvanih konvencionalnih prijevoda (engl. *human translation*) i automatiziranih strojnih prijevoda koji nastaju u računalnim programima bez izravne ljudske intervencije (engl. *machine translation*), a razvojnim je programerima osobito važno vrednovati izlazne podatke sustava za strojno prevođenje kako bi utvrdili koje se njihove komponente još moraju unaprijediti da bi rezultati bili što bolji. Stoga se vrednovanje dosljedno provodi u svim fazama razvoja sustava za strojno prevođenje, učestalo upravo ljudskom procjenom, i to unatoč tome što su takvi postupci dugotrajni, skupocjeni i u svojoj naravi subjektivni (Girardi i sur. 2014: 120).

U uskoro šezdeset godina dujoj povijesti vrednovanja strojnog prevođenja, čiji se početak smješta u 1966. godinu kada je objavljen poznati ALPAC-ov izvještaj (v. potpoglavlje 2.1), osmišljene su brojne metode koje razvojnim programerima, prevoditeljima i ciljnim korisnicima pomažu utvrditi kvalitetu dobivenih izlaznih podataka. Neke od njih danas se više ne upotrebljavaju, dok se druge unaprjeđuju i tek doživljavaju svoj procvat. No usprkos dugogodišnjim naporima istraživača do današnjeg dana nije osmišljena standardna procedura vrednovanja strojnog prevođenja koja bi bila sveopće prihvaćena (Freitag i sur. 2021: 1460). Teoretičari i profesionalni prevoditelji zapravo se „uvelike slažu da nema jedinstvenog objektivnog načina za mjerenje kvalitete“¹ (Drugan 2013: 35), a okosnica ove disertacije

¹ Sve citate inozemnih autora upotrijebljene u ovom doktorskom radu prevela je doktorandica, osim ako nije drugačije naznačeno.

upravo su različiti načini na koje čovjek može vrednovati strojne prijevode, pri čemu će se odabrane metode razložiti na njihove sastavnice, uz promjene varijabli i procjene prikladnosti različitih modela vrednovanja. Očekivani praktični ishodi ovog rada uključuju prikupljanje vrijednih saznanja o ljudskim načinima vrednovanja strojnog prevođenja, pogotovo u kontekstu sve boljih rezultata koji se postižu najnovijim tehnologijama strojnog prevođenja, čak i kada se prevodi na jezike koji imaju razmjerno manje digitalnih resursa, kao što je i hrvatski. Svako istraživanje o strojnom prevođenju provedeno unutar discipline računalne lingvistike u Hrvatskoj podiže svijest o mogućnostima i ograničenjima alata za strojno prevođenje u danom trenutku, a u okviru ovog istraživanja dobit će se uvid u to kako trenutni stavovi o strojnom prevođenju utječu na vrednovanje strojnih prijevoda te kako se pri vrednovanju mogu postići relevantniji i objektivniji podaci. Istraživanja strojnog prevođenja također su usko povezana s područjem znanosti o prevođenju te su očekivani ishodi ovog istraživanja korisni i za praktičnu prevoditeljsku svakodnevicu, s obzirom na to da razvijanje alata za strojno prevođenje direktno utječe na promjene u njihovoj struci i radu te je važno da prevoditelji na raspolaganju imaju što točnije i ažuriranije podatke o stanju strojnog prevođenja na hrvatski jezik.

U drugom poglavlju izložen je teorijski okvir rada, počevši s kratkom poviješću strojnog prevođenja, s osobitim naglaskom na pristupe strojnom prevođenju kroz povijest, od prvih sustava koji su se temeljili na pravilima do najsuvremenije tehnologije neuronskog strojnog prevođenja, a potom su opisani i prisutni strahovi dijela prevoditelja pred rapidnim razvojem prijevodnih tehnologija te mitovima koji o strojnom prevođenju kruže od početka istraživanja ovog tipa. Potpoglavlje koje tomu slijedi posvećeno je problematici i kriterijima vrednovanja strojnih prijevoda, što ljudskim metodama, što automatskim metrikama, a potom su potanko predstavljene i istaknutije metode vrednovanja: metoda subjektivne procjene (s opisanim glavnim problemom niskih razina slaganja među ispitanicima te mogućim kriterijima i bodovnim ljestvicama za provedbu metode), metoda provjere razumljivosti testovima razumijevanja (uz definiciju pojma razumljivosti, prikaz tehnika provjere razumljivosti i praksi pripreme testova razumijevanja), i metoda analize pogrešaka (s opisanim izazovima, ali i uspješnim načinima provedbe analize, te prikazom mogućih kategorija pogrešaka i njihova odnosa s percepcijom kvalitete; unutar tog potpoglavlja na koncu je predstavljena i povezana metoda redakture strojnih prijevoda).

U trećem poglavlju predstavljen je cilj ovog istraživanja te je na temelju dosadašnje literature i spoznaja o vrednovanju strojnog prevođenja pojašnjeno kako su izniknula istraživačka pitanja koja su predmet ovog istraživanja.

Četvrto poglavlje posvećeno je metodologiji postavljenog istraživanja koje se sastoji od nekoliko manjih zadataka, koji su svi zasebno opisani, a uključuju subjektivnu procjenu kvalitete strojnog prijevoda, test razumijevanja, anketni obrazac o stavovima i iskustvima ispitanika te analizu pogrešaka. Nadalje se opisuje materijal na kojemu je istraživanje provedeno, broj i vrsta ispitanika koji su odabrani za sudjelovanje u ovom istraživanju te sama metoda provedbe ovog kvalitativnog istraživanja.

Peto poglavlje namijenjeno je prikazu rezultata istraživanja te diskusiji, grupirano po pojedinim metodama i istraživačkim pitanjima: u potpoglavlju 5.1 iznose se rezultati i raspravlja o dvama istraživačkim pitanjima koja se odnose na metodu subjektivne procjene; potpoglavlje 5.2 odnosi se na rezultate i diskusiju o metodi testa razumijevanja; u potpoglavlju 5.3 prikazani su rezultati analize pogrešaka te se raspravlja o sljedećim dvama istraživačkim pitanjima; potpoglavlje 5.4 posvećeno je rezultatima i raspravi o dijakronijskoj usporedbi metoda u odnosu na opsežnije preliminarno istraživanje provedeno 2019. godine; a u potpoglavlju 5.5 donosi se završna diskusija o svim uključenim metodama vrednovanja i doprinosima ovog istraživanja.

Šesto i zadnje poglavlje zaključak je donesen na temelju svih dobivenih rezultata, provedene diskusije te teorijskog pregleda.

Nakon popisa literature i sažetka na hrvatskom i engleskom jeziku slijede prilozi u kojima su na raspolaganje stavljeni: usporedni prikaz testnog materijala odnosno izvornika, dvaju verzija izlaznih podataka (iz 2019. i 2021. godine) te referentnog prijevoda (Prilog A), cjelokupni anketni obrazac dostupan za slobodno korištenje (Prilog B), tablica s razinama slaganja ispitanika s određenim izjavama kao i kodirani stavovi ispitanika koji su skupno i pojedinačno komentirani u radu (Prilog C), puna tablica s rezultatima testa razumijevanja iz 2019. i 2021. godine čiji su dijelovi pobliže tematizirani u radu (Prilog D) te puni usporedni prikaz ocjena ispitanika za pročitani tekst, broj točnih odgovora na testu razumijevanja i kontrolno pitanje sigurnosti u odabire na testu (Prilog E). Prilozi, zajedno s testnim materijalom koji čine izvorni njemački tekst, referentni ljudski prijevod izvornika te strojni prijevod napravljen u alatu za strojno prevođenje, a koji su predstavljeni u potpoglavlju 4.1 (Materijal), omogućuju replikabilnost i provjeru ovdje prikazanih rezultata, a testni set u budućim

istraživanjima strojnog prevođenja može, uz potrebne izmjene, poslužiti i za ispitivanje posve novih hipoteza.

2. Teorijski okvir

Vještina prevoditelja koji prevode tekstove s izvornoga na ciljni jezik daleko nadmašuje puko poznavanju tih dvaju jezika. Prevođenje je složen kognitivni, lingvistički, društveni, kulturalni i tehnološki proces, što i mjerenje njegove kvalitete čini složenim procesom (Castilho i sur. 2018: 10). Upravo je stoga kroz povijest i predstavljalo velik problem kako računala podučiti prevođenju. Neki od čimbenika koji tomu pridonose jesu: velike varijacije u definiranju prijevodne ekvivalencije ovisno o vrsti teksta koji se prevodi, strukturne međujezične razlike, leksičke i terminološke praznine u ciljnom u odnosu na izvorni jezik te veliki zahtjevi za kreativnošću kod prilagodbe teksta drugoj kulturi (Arnold 2003: 119-120). Računalima najveće poteškoće otpočetak zadaju dvosmislenost, strukturne i leksičke razlike među jezicima te ustaljene višerječne jedinice poput frazema i kolokacija (Arnold i sur. 1994: 105).

Važno je, međutim, naglasiti da je osobito u zadnjem desetljeću došlo do skokovitih pomaka u kvaliteti strojnih prijevoda. Tehnologija je, unatoč svim navedenim poteškoćama, ipak drastično unaprijeđena. To, doduše, ne prati odmah i napuštanje brojnih mitova o strojnom prevođenju koji su posljednjih desetljeća niknuli iz plodnog tla neuspješnih prijevoda i pogrešaka koje su tijekom godina zabilježene u strojnim prijevodima. Teorije koje su danas svakako, a možda i u vrijeme njihova nastanka bile neutemeljene, kao i predrasude o strojnom prevođenju najprije valja raščistiti, i treba smjestiti takve misli u povijesni kontekst iz kojeg su potekle (v. potpoglavlje 2.1). U drugom dijelu ove cjeline donosi se pregled najvažnijih saznanja o vrednovanju strojnog prevođenja (v. potpoglavlje 2.2), a u trećem su razložene i pojedine metode (v. potpoglavlje 2.3).

2.1 Kratka povijest strojnog prevođenja

Prvotna se zamisao o strojnom prevođenju pripisuje američkom matematičaru Warrenu Weaveru koji je, zacijelo inspiriran razbijanjem neprijateljskih kodova u Drugom svjetskom ratu, zamislio da je ruski tekst pred njim zapravo engleski tekst zapisan nekim neobičnim simbolima i da će zapisanim informacijama moći pristupiti ako jednostavno dešifrira kod (Arnold i sur. 1994: 12). Njegovim je zapisom, poslije poznatim kao Weaverovim memorandumom, u Americi započelo doba intenzivnog istraživanja i ulaganja u razvoj sustava za strojno prevođenje koje je trajalo petnaest godina, sve dok nije osnovan odbor ALPAC

(*Automated Language Processing Advisory Committee*) čiji je zadatak bio procijeniti dotadašnji napredak u području strojnoga prevođenja – i koji je naglo dokinuo svaku vjeru u strojno prevođenje kao ostvarivu mogućnost. Odbor je izvijestio da na engleskom govornom području prvenstveno uopće ne nedostaje profesionalnih prevoditelja, dapače, da ih je daleko više nego što za njima postoji potreba (ALPAC 1966: 11) te da se o razvoju strojnog prevođenja zapravo uopće ne može govoriti jer ne postoji sustav koji postiže ikakve korisne rezultate niti se nazire mogućnost da bi se i u kakvoj bližoj budućnosti takav sustav mogao razviti (isto, 32). Kao dokaz tomu odbor je naveo i da je Sveučilište u Georgetownu i poslije osam godina istraživanja strojnog prevođenja i dalje dobivalo samo izlazne podatke koje je trebalo redigirati, što je i trajalo duže i bilo skuplje od konvencionalnog načina prevođenja (isto, 19). ALPAC-ovo je izvješće razorno utjecalo na daljnji razvoj strojnog prevođenja, ali u njemu se pak s druge strane prvi put ozbiljno pristupilo razvoju objektivne metodologije za vrednovanje strojnih prijevoda.

Ostvarenje potpuno automatiziranog strojnog prevođenja (engl. *fully automatic high quality machine translation*), koje se tada uzimalo kao glavni cilj istraživanja, poslije su zamijenile težnje prema *dovoljno dobrim* prijevodima koji bi bili korisni za određene uporabe i kod kojih se prihvaća potreba za doradivanjem, pa se već desetak godina poslije ponovno javio interes za automatiziranim prevođenjem. U drugom su valu istraživanja sustavi za strojno prevođenje izišli iz laboratorija i postavili se na tržište, a ovaj su put i očekivanja bila znatno realističnija (Dovedan, Seljan i Vučković 2002: 286): tako se, primjerice, kanadska vlada 1976. godine počela služiti sustavom METEO koji se pokazao korisnim za prijevode svakodnevnih vremenskih izvješća s engleskoga na francuski jezik, a on je u uporabi još i danas; Europska zajednica preuzela je rusko-englesku inačicu sustava SYSTRAN, razvijenu na već spomenutom Sveučilištu u Georgetownu, i počela je nadograđivati u višejezični sustav; a tada su se stali pojavljivati i sustavi s drugim jezičnim kombinacijama, poput SPANAM-a i ENGSPAN-a za prevođenje između španjolskog i engleskog jezika, sustavi koji prevode s japanskog i sl. Razvijajući i unaprjeđujući sustave za strojno prevođenje, lingvisti, ali i brojni fizičari i programeri, osmislili su različite pristupe strojnom prevođenju.

2.1.1 Pristupi strojnom prevođenju kroz povijest

Prvi sustavi za strojno prevođenje temeljili su se na pravilima (engl. *rule-based*). Oni funkcioniraju tako da lingvističkom analizom raščlane rečenicu na njezine dijelove, zatim u ugrađenim rječnicima potraže prijevodne ekvivalente za svaki rečenični dio i na koncu prema naučenim pravilima preurede redosljed riječi tako da odgovara ciljnom jeziku. Tri su generacije sustava za strojno prevođenje koji se temelje na pravilima (Tripathi i Sarkhel 2011: 389): izravni sustavi, transferni sustavi (u kojima se rečenica na izvornom jeziku prvo predoči u apstraktnom obliku, a zatim se u drugom koraku izražava na ciljnom jeziku) i sustavi koji rade s međujezikom (engl. *interlingua*, tako nazvanih jer se rečenica na izvornom jeziku više ne izražava apstraktnim jezikom, nego nekim trećim prirodnim jezikom, a potom se iz tog međujezika prelazi u ciljni). Nedostaci prvih, pretežno lingvistički orijentiranih sustava brojni su: rječnici nisu dovoljno opsežni za te potrebe, ne obuhvaćaju sve pojavnice i nema ih dovoljno, dio jezičnih informacija mora se unositi ručno, velik broj pravila međusobno se sukobljava, sustavi se loše snalaze s dvosmislenostima i frazemima, a takve je sustave i vrlo skupo hraniti novim leksikom i tako ih osposobljavati za prevođenje novih područja (Okpor 2014: 161).

U drugoj su se fazi razvoja sustava za strojno prevođenje iskušali empirijski sustavi, temeljeni na podacima (engl. *knowledge-based*). Najveći interes uživali su statistički sustavi (u kojima se ekvivalenti traže primjenom statističkih metoda) predvođeni skupinom fizičara zaposlenih u tvrtki IBM koji su prevođenju pristupili kao problemu vjerojatnosti te su, „umjesto definiranja gramatičkih pravila, stvorili program koji se koristio stvarno prevedenim tekstovima“, odnosno „računalo je jednostavno nizalo riječi u različitim kombinacijama dok ne bi dobilo statistički najvjerojatniji slijed“ (Simeon 2008: 18). Na tom tragu razvijani su i sustavi koji se temelje na primjerima, a oni svoju primjenu u prvom redu pronalaze u okviru strojno potpomognutog prevođenja², stvarajući prevoditeljske memorije kojima se prevoditelji služe kako bi „ekstrahirali gotove prijevode riječi, kolokacija, pa čak i čitavih rečenica“ (isto, 20). Okosnicu ovih dvaju pristupa čine usporedni korpusi. No i statistički su sustavi kritizirani zbog

² Strojno potpomognuto prevođenje odnosi se na situaciju kada čovjek (a ne računalni program) prevodi služeći se kao pomoći specijaliziranim računalnim programima poput: sustava prijevodnih memorija, sustava za upravljanje terminologijom, alata za lokalizaciju mrežnih stranica, alata za lokalizaciju softvera, alata za podslovljavanje, alata za osiguranje kvalitete (Pavlović 2015: 292). S druge strane, „u slučaju strojnog prevođenja, računalni program sam prevodi tekst, a čovjek se uključuje naknadno, u fazi redakture teksta“ (isto, 25).

svojih nedostataka, osobito zato što znatno slabije funkcioniraju između jezika čija se pravila za redoslijed riječi u rečenici uvelike razlikuju (npr. između japanskog i engleskog jezika), ali i zbog visokih troškova stvaranja dovoljno velikih usporednih korpusa (Okpor 2014: 163). Statistički su sustavi na vrhuncu bili između 2004. i 2014. godine (Kenny 2022: 39), kada palicu preuzimaju neuronski sustavi.

2.1.2 Neuronsko strojno prevođenje

Neuronsko strojno prevođenje (engl. *neural machine translation*) nova je paradigma koja se posljednjih godina našla u žarištu istraživanja. Temelji se na neuronskim mrežama i velikim skupovima podataka kroz koje sustav sam „uči“ prevoditi. Naša znanja o tome kako neuronsko strojno prevođenje to točno čini još uvijek su ograničena (Tan i sur. 2020: 18). Neuronske mreže simplificirano se može shvatiti kao iznimno kompleksnu matematičku formulu čiji su ishod prijevodi *sekvenci* (što mogu biti čitavi dokumenti, odlomci ili pojedine rečenice), što se postiže prvo kodiranjem izvorne rečenice u kodove koji predstavljaju pojedine riječi izvornog jezika, a zatim dekodiranjem tih kodova u riječi ciljnog jezika, služeći se algoritmima koji su na engleskom jeziku poznati pod nazivima *back-propagation*, *greedy-search* ili *beam-search* da bi se postigli najbolji rezultati (usp. Tan i sur. 2020 za detaljan matematički prikaz formula u pozadini djelovanja neuronskih mreža i neuronskog strojnog prevođenja). Neuronsko strojno prevođenje na osnovnoj razini funkcionira, dakle, u dvjema fazama: u fazi kodiranja svakoj se riječi pridružuje tzv. „neuralna reprezentacija“ (engl. *neural representation* ili *embedding*) koje se onda spajaju na razini rečenice i tada prilagođavaju ovisno o kontekstu; u fazi dekodiranja te se reprezentacije na razini rečenice raspliću kako bi predvidjele prijevod pojedinih riječi (Pérez-Ortiz, Forcada i Sánchez-Martínez 2022: 143). Kao i uvijek kod uključivanja umjetne inteligencije, koja se često definira kao „disciplina računalne znanosti koja stvara strojeve – odnosno računalne programe – koji mogu rješavati probleme kakvi obično zahtijevaju ljudsku inteligenciju“ (Kenny 2022: 34-35), nije potrebno da računalni programi nužno *razmišljaju* kao ljudi, već samo da *djeluju* kao inteligentni ljudi.

Dorothy Kenny neuronske sustave pojašnjava kroz računalne *modele*, koji su apstraktne, matematičke reprezentacije realnih događanja, sustava ili fenomena, a čija je uloga da predvide odgovor na dotad neviđeni problem, kao što je prijevod nekog dosad neprevedenog segmenta (2022: 40). Kod neuronskih su sustava ti modeli nadahnuti radom ljudskog mozga: služe se

umjetnim neuronskim mrežama s tisućama individualnih jedinica – umjetnih neurona – koji su povezani s tisućama drugih umjetnih neurona, koji se aktiviraju kada primaju informacije od drugih neurona (isto). Da bi jedna takva mreža neurona mogla „učiti“ prevoditi dotad neviđene kombinacije jezičnih elemenata, važno je da reprezentacije sličnih elemenata budu slične. Juan Antonio Pérez-Ortiz, Mikel L. Forcada i Felipe Sánchez-Martínez daju primjer engleskih riječi: *poured*, *rained*, *pouring* i *raining* – ove bi četiri riječi u idealnom slučaju bile slično kodirane jer pripadaju istom semantičkom polju, s tim da bi riječi *poured* i *rained* trebale biti sličnije kodirane jer su oboje izražene prošlim glagolskim vremenom, dok bi *pouring* i *raining* pak morale biti kodirane sličnije drugim riječima koje su izražene gramatičkom kategorijom gerundiva, npr. *driving*, jer je vjerojatno da će se takve riječi naći u sličnom kontekstu (2022: 152). To se pridruživanje kodova riječima naziva dubokim učenjem (engl. *deep learning*) koje neuronskim sustavima pomaže da što bolje prevode (Koehn 2020: 64). Riječ je, dakle, o iznimno kompleksnim mrežama umjetnih neurona, koje su ovdje prikazane samo u najjednostavnijim crtama, dok se više tehničkih detalja potrebnih za dubinsko razumijevanje neuronskih sustava za strojno prevođenje može pročitati u monografiji „Neural machine translation“ Philippa Koehna (2020) u kojoj autor detaljno opisuje funkcioniranje neuronskih mreža, neuronskih modela, kodiranja i dekodiranja te načela treniranja sustava.

Očekivanja od neuronskog prevođenja otpočetak su vrlo visoka i brzo od njegova uvođenja uslijedila su brojna istraživanja u kojima se provjeravaju izlazni podaci neuronskih sustava. Luisa Bentivogli i sur. (2016: 265) usporedili su rad jednog neuronskog i triju statističkih sustava, i kod neuronskog sustava zamijetili veliki napredak: u redakturu njegovih izlaznih podataka uloženo je 26 % manje truda u odnosu na najbolje rangirani statistički sustav; neuronski sustav proizveo je bolje prijevode u svim rečenicama osim u onim najduljima; neuronski sustav postigao je dobre rezultate i u tekstovima s bogatim leksikom; te su izlazni podaci neuronskog sustava, u odnosu na najbolje rangirani statistički sustav, sadržavali 19 % manje morfoloških i 17 % manje leksičkih pogrešaka, kao i 50 % manje pogrešaka u redosljedu riječi (od čega se ističe čak 70 % manje pogrešaka u redosljedu glagola u rečenici).

U istraživanju koje su proveli Sheila Castilho i sur. rezultati usporedbe neuronskih i statističkih sustava za različite jezične parove nisu bili dosljedni što se tiče točnosti i truda koje je potrebno uložiti za redakturu teksta, ali su izlazni podaci neuronskog sustava dokazano bili tečniji (2017: 109), te su imali veći broj prevedenih segmenata u kojima se nije javila nijedna pogreška (isto, 116). S druge strane, na svim je jezicima uočeno više pogrešno izostavljenih i

umetnutih riječi te su zabilježena i potpuno neadekvatna prijevodna rješenja (isto, 118). Slično zamjećuju i Yonghui Wu i sur. (2016: 2): u njihovu se istraživanju neuronski sustav nije pokazao dovoljno robusnim za prijevode riječi koje se rjeđe pojavljuju i u izlaznim su podacima katkad nedostajali dijelovi izvornih rečenica, odnosno nisu bile prevedene sve riječi.

Antonio Toral i Víctor M. Sánchez-Cartagena potvrdili su da su izlazni podaci neuronskih modela tečniji, da je redosljed riječi u prevedenim rečenicama točniji i da neuronski sustavi imaju manje poteškoća sa sklanjanjem (2017: 1071), a kao nedostatak izdvojili su pad kvalitete kod rečenica koje su dulje od 36 do 40 riječi (isto, 1069). Generalno se smatra da manji broj pogrešaka i tečniji izlazni podaci olakšavaju redakturu prijevoda (Klubička, Toral i Sánchez-Cartagena 2018: 213), ali iznimno tečni rezultati katkad mogu biti i varljivi – neke besprijekorno tečne ciljne rečenice zapravo ne moraju sasvim odgovarati izvornim rečenicama (Way 2020: 320).

U obzir se mora uzeti i da istraživači najčešće rade s engleskim jezikom, a da strojno prevođenje postiže slabije rezultate u radu s jezicima za koje je dostupno manje resursa, premda to ne odgovara načinu kako se ljudi diljem svijeta služe (strojnim) prijevodima te dovodi do lošijih rezultata prevođenja između jezika koji nisu engleski (Fan i sur. 2021: 2). Usporedbom izlaznih podataka koji su sa švedskog na hrvatski jezik najprije prevedeni statističkim sustavom, a zatim recentno uvedenom neuronskom inačicom istog sustava, pokazalo se da izlazni podaci u svakom slučaju vrve leksičkim i grubim morfosintaktičkim pogreškama, a nerijetko i potpuno nejasnim rečenicama (Ljubas 2018: 89), iako su izlazni podaci neuronskog sustava i u tom slučaju bili tečniji i uspješniji od izlaznih podataka statističkog sustava.

Ponukani sve boljim izlaznim podacima neuronskih sustava Guerberof-Arenas i Toral proveli su istraživanje o kreativnosti u konvencionalnom i strojnom prevođenju, i to na književnom tekstu kao predlošku, no ispitanici su se jednoglasno složili da je recenzirani ljudski prijevod „iznimno dobar“, strojni prijevod „iznimno loš“, a strojni prijevod koji je naknadno uređen „ni dobar ni loš“ (2022: 14). Ispitujući stavove prevoditelja koji su sudjelovali u istraživanju, autori su također zaključili da prevoditelji smatraju da ih predložak strojnog prijevoda pri redigiranju ograničava, uspoređujući ga s „nošenjem korzeta“ (isto, 22), a analizom teksta autori zaključuju da i neuronski strojni sustavi „trenutno nemaju potrebne kapacitete za kreativne prijevode; pružaju doslovna rješenja za prijevodne probleme“. Waltraud Kolb također dolazi do tog zaključka, no navodi druge oblike korištenja strojnog prevođenja u svijetu književnog prevođenja – izdavači, primjerice, uz dobre rezultate strojnog prevođenja

mogu stvoriti prvi dojam o nekoj knjizi koju bi potom ponudili profesionalnom prevoditelju na prijevod (2022: 20). S prednostima i nedostacima neuronskog strojnog prevođenja još se upoznajemo, no o velikoj vjeri u njega dovoljno govori činjenica da su i najveći pružatelji besplatnih usluga prevođenja, primjerice Google ili Bing, prešli na neuronske modele prevođenja.

2.1.3 Mitovi, strahovi i predrasude o strojnom prevođenju

Pristupi su se mijenjali i različiti sustavi razvijali, ali kroz čitavo se vrijeme razvoja ovog područja u računalnoj lingvistici za njega zadržao naziv „strojno prevođenje“ (engl. *machine translation*), premda bi danas svakako bilo prikladnije govoriti o automatiziranom prevođenju. Računala, naime, već odavno ne percipiramo kao nekakve misteriozne „strojeve“ (Somers 2012: 284), nego ih smatramo ključnim prevoditeljskim alatima, a opseg znanja i vještina povezanih s prevoditeljskim tehnologijama kojima prevoditelji budućnosti moraju baratati također raste iz dana u dan (Kučiš i Seljan 2014: 304). Istraživanje provedeno s dvjema skupinama studenata, od kojih se jedna služila internetskim korpusima i višejezičnim terminološkim bazama, a druga je prevodila bez njih, pokazalo je da su prijevodni prevoditelja koji se takvim alatima služe daleko bolji, uzevši u obzir broj leksičkih, pravopisnih i interpunkcijskih pogrešaka (Kučiš, Seljan i Klasnić 2009: 349). Računala nam, dakle, pomažu i nisu nam strana. Kako bi se smanjile predrasude o strojnom prevođenju i strahovi koje prevoditelji nerijetko imaju, a koji se najčešće odnose na to da bi ih strojevi mogli zamijeniti ili im „ukrasti“ posao, važno je demistificirati prevoditeljske tehnologije povezane s računalima, integrirati ih u proces obrazovanja prevoditelja i smatrati ih dijelom stvarnosti koju uvjetuju globalizacija i potreba za brzim protokom informacija (Kučiš i Seljan 2014: 304). Među te se tehnologije, osim sustava za strojno prevođenje, ubrajaju i mrežni rječnici, terminološke baze, korpusi, alati za izvlačenje terminologije te prijevodne memorije.

Douglas Arnold i sur. donose detaljan pregled najranijih mitova, od kojih su neki i danas još vrlo popularni: primjerice, iz davnina se može čuti da je strojno prevođenje gubitak vremena jer nikada neće postojati sustav koji bi znao prevesti Shakespeareova djela, ili da je posve beskorisno zato što je kvaliteta izlaznih podataka sustava za strojno prevođenje vrlo loša (1994: 7). Istovremeno se – i proturječno – otpočetka govori o tome kako strojno prevođenje prevoditeljima ugrožava egzistenciju, te rasprave o strojnom prevođenju započinju i završavaju

pitanjem hoće li „strojevi“ zamijeniti ljude. Utvrđeno je da prevoditelji redigirajući strojne prijevode unose višestruke suvišne ispravke u već prihvatljive prijevode, samo zato što bi oni osobno nešto izrazili na drugi način (a sličan trend zamjećuje se i kada lektori uređuju konvencionalne prijevode) – uostalom i zato što žele opravdati svoj položaj u prevoditeljskom procesu (Way 2018: 166). Zbog tog strahovanja za egzistenciju prevoditelji često nisu svjesni da se uvođenjem sustava za strojno prevođenje zapravo može poboljšati njihovo zadovoljstvo na poslu jer su sustavi za strojno prevođenje najučinkovitiji u zamornim i repetitivnim zadacima. U novije se vrijeme, međutim, strahovi prevoditelja od sustava za strojno prevođenje mogu tražiti i u odnosu poslodavaca prema njima: uvođenje alata za strojno prevođenje neprestano smanjuje honorare u određenim sektorima prevođenja, te se od prevoditelja očekuje da rade brže za manje plaće, što kod prevoditelja pak stvara potrebu da se diferenciraju od običnog alata (O'Brien 2022b: 72).

Istraživači su godinama jasno naglašavali da se ne mogu svi tekstovi strojno prevesti tako da rezultati budu prihvatljivi (Bennett i Gerber 2003: 186), pa je stoga bespredmetno bilo strojnom prevođenju podmetati prevođenje Shakespeareovih soneta i općenito lijepe književnosti. Istraživanja strojnog prevođenja zasad su prvenstveno³ usmjerena na tekstove čije su značajke izravnost i nedvosmislenost, a to su, primjerice, tehnički tekstovi, stručna izdanja ili općenito poslovna komunikacija (isto). Strojno prevođenje također pospješuje čitanje mrežnih stranica i članaka pisanih na stranom jeziku, pružanje korisničke podrške ili osobnu komunikaciju na društvenim mrežama.

Kada govorimo o prihvatljivosti rezultata, važno je osvijestiti da nisu svi strojni prijevodi namijenjeni objavljivanju. S konvencionalnim se prijevodima nikada nije javilo pitanje namjene „polu-dovršenih“ prijevoda jer profesionalni prevoditelji takve prijevode načelno ne isporučuju (Allen 2013: 298), ali danas su čak i takvi prijevodi praktični za internu ili privatnu uporabu. S jedne se strane govori o „diseminaciji“ podataka, pri čemu se misli na strojne prijevode kojima se profesionalni prevoditelji služe kako bi na njima temeljili konačne prijevode, odnosno kojima se služe kao prvotnim nacrtima, a s druge strane o „asimilaciji“ podataka, u situacijama kada se bilo tko može poslužiti strojnim prijevodom kako bi razumio suštinsko značenje nekog teksta (Ghyselen 2017: 1). Jasno je da osjetljivi dokumenti, osobito

³ Premda u najrecentnije vrijeme ne i isključivo, v. rezultate istraživanja Guerberof-Arenas i Toral (2022) prikazano u poglavlju 2.1.2., koje je dio šireg projekta Europske unije pod nazivom *Creativity and narrative engagement of literary texts translated by translators and neural machine translation*, a koji se bavi upravo proučavanjem uporabe neuronskog strojnog prevođenja na kreativnijim prevoditeljskim zadacima.

oni koji sadrže informacije za sigurnost korisnika ili koji su važni u društveno-političkom smislu, zahtijevaju pažljivije izrađene prijevode, no ljudi sve češće imaju potrebu razumjeti samo osnovne misli izražene u nekom tekstu koji nije napisan na njihovu materinskom jeziku ili nekom stranom jeziku koji govore – a uz pomoć sustava za strojno prevođenje, osobito onih javno dostupnih na internetu, to danas vrlo lako i mogu (Allen 2013: 300).

O kvaliteti takvih prijevoda može se raspravljati, no milijuni se ljudi ipak svakodnevno služe sustavima za strojno prevođenje, a to govori da su korisnici zacijelo barem u nekoj mjeri zadovoljni razinom kvalitete izlaznih podataka, inače bi se prestali njima služiti (Way 2018: 159). Mnogi su takvi alati besplatni te lako dostupni širokom broju korisnika. Tako je, primjerice, Google Prevoditelj, alat koji je među prevoditeljima u mnogim zgodama predmet ismijavanja zbog pogrešaka koje se u njegovim prijevodima javljaju, još 2012. godine brojio više od dvjesto milijuna aktivnih korisnika te je iste godine u samo jednom danu prevodio onoliko količinu teksta koliku bi svi prevoditelji svijeta preveli u godinu dana (Och 2012), dok se u trenutku pisanja ovog rada broj svakodnevnih korisnika procjenjuje na više od šesto milijuna. S tim na umu, teško je govoriti o beskorisnosti i neupotrebljivosti strojnog prevođenja, koliko god daleko od savršenih oni bili. U zadnje se vrijeme također govori o potrebi za novom vrstom pismenosti, kao što se tijekom recentnih desetljeća razvila potreba i za medijskom pismenošću, računalnom ili informacijskom pismenošću, pa je tako možda danas već potrebno i da svi korisnici kojima su sustavi za strojno prevođenje dostupni imaju neku osnovnu razinu poznavanja njihova djelovanja: laici će i sami vrlo lako naučiti kopirati željeni tekst u za to predviđenu kućicu i uživati u prevedenim izlaznim podacima, ali potrebna su i druga znanja da bi se strojnim prijevodima pristupilo kritički ili da bi se od njih izvuklo najviše koristi (Bowker i Buitrago-Ciro 2019: 33). S druge bi strane u kontekstu akademske komunikacije ta predložena pismenost, između ostalog, uključivala razumijevanje osnova o tome kako sustavi za strojno prevođenje obrađuju tekstove, znanje o tome kako pisati tekstove da se lakše mogu strojno prevesti ili znanje o tome kako uređivati izlazne podatke da bi se pospješila njihova točnost (isto, 88).

Naime, ako su strojni prijevodi i namijenjeni objavljivanju, onda se oni ipak najčešće ne objavljuju kao sirovi strojni prijevodi, već se uređuju. Redaktura prijevoda koji je s izvornog na ciljni jezik preveden sustavom za strojno prevođenje (engl. *post-editing*) provodi se kako bi prijevod udovoljio zahtjevima kvalitete kakvu moraju imati konvencionalno i profesionalno prevedeni tekstovi, a to uključuje uređivanje, izmjenjivanje i/ili ispravljanje teksta (Allen

2003: 297). Danas je postupak redakture strojnih prijevoda svakodnevnica velikog (i sve većeg) broja profesionalnih prevoditelja u svijetu, pa tako i u Hrvatskoj, čak i ako mnogi od njih smatraju da je takav postupak dugotrajniji od samoga prevođenja (Koletnik Korošec 2011).

Uvođenje sustava za strojno prevođenje neupitno donosi promjene u radnoj svakodnevici, ali prevoditelji ih svejedno ne moraju sumnjičiti kao „dehumanizirajućih čudovišta“ koja im dolaze preotimati poslove jer njihova svrha nije zamijeniti ljude na radnim mjestima, već olakšati prevoditeljski posao koji nije uvijek samo kreativan i lagodan, već zahtijeva vrijeme, trud i donošenje mnogih teških odluka (Schäfer 2003: 134). Kada se prevoditelj služi strojnim prijevodom kao nacrtom za budući prijevod, njegov se zadatak sastoji od provjeravanja rečenice po rečenicu prijevoda u odnosu na izvorni tekst te utvrđivanja i ispravljanja svih pogrešaka koje je sustav za strojno prevođenje napravio (isto, 135). Kako bi kvalitetno ispunio taj zadatak, od prevoditelja se očekuje visoka razina stručnosti, ali ga se treba i obučiti za taj posao, primjerice, očekuje se da prevoditelj nauči prepoznati tipične „strojne“ pogreške.

Činjenica da sustavi za strojno prevođenje mnogo lošije rade kada prevode na takozvane „jezike s malo računalnih izvora“ u Hrvatskoj zacijelo pridonosi predrasudama prema strojnom prevođenju i navedenim promjenama u prevoditeljskom radu. Hutchins (2012: 298) objašnjava da nije bitno koliko govornika nekog jezika ima u svijetu, već za koje jezike nema osobito mnogo dostupnih jezičnih resursa ili koji se ne smatraju komercijalno ili gospodarski značajnima. U te se jezike ubraja i hrvatski (Somers 2003: 90). Za strojno su prevođenje pogodnije one jezične kombinacije „na kojima se puno radilo i/ili koje raspoložu velikim brojem usporednih korpusa, a prijevodi između srodnih jezika u pravilu su bolje kvalitete“ (Pavlović 2017: 280). Sustavno ulaganje u razvoj jezičnih tehnologija za hrvatski jezik ključno je ako u budućnosti želimo vidjeti kvalitetnije strojne prijevode na hrvatski jezik, osobito u kombinaciji s drugim manjinskim jezicima (Ljubas 2017: 60), a neke su od najvažnijih tehnologija koje nedostaju jezicima u manjini sljedeće: programi za obradu riječi, provjernici pravopisa, jednojezični i dvojezični rječnici, korpusi i sl. (Hutchins 2012: 298). Sve to hrvatskom jeziku u dovoljnoj mjeri nedostaje još i danas, u vrijeme pisanja ovog rada. Dugoročno gledano, istraživače brine mogućnost da će nedostatak resursa dovesti do osiromašenja takvih jezika, osobito ako će sustavi za strojno prevođenje zbog nedostatka svježeg dotoka konvencionalno prevedenih materijala imati sve više problema s držanjem koraka sa suvremenim jezikom (Moorkens 2022: 135).

Takva i slična pitanja razmatraju se u okviru etike strojnog prevođenja, koja polazi od filozofske misli da nijedna tehnologija nije etički neutralna, već uvijek zrcali vrijednosti proizvođača tehnologije. Polazeći od te premise, Joss Moorkens u prilogu „Ethics and machine translation“ tematizira etička pitanja povezana s današnjom tehnologijom strojnog prevođenja, primjerice, navodi kako korisnici javno dostupnih sustava za strojno prevođenje upisujući tekstove koje žele prevesti ujedno i nesvjesno ustupaju pravo na korištenje tih podataka, čime katkad razotkrivaju povjerljive i osjetljive podatke (2022: 131). Također, podsjeća da je treniranje sustava za neuronsko strojno prevođenje ujedno i ekološko pitanje jer se u tom procesu proizvede gotovo pet puta više emisije CO₂ nego što prosječan automobil proizvede za svog vijeka, a i čitava je industrija informacijske i komunikacijske tehnologije ovisna o rudarenju rijetkih metala te poznata po slabim praksama recikliranja i većem zagađenju okoliša (isto, 133-134). Etička su to pitanja koja nadilaze prevoditeljske prakse, ali ukazuju na mnoga utemeljena razmišljanja o negativnim implikacijama snažnog širenja strojnog prevođenja.

Osnovna je pak briga prevoditelja, osobito slobodnih prevoditelja, to što se njihovi prijevodi upotrebljavaju za treniranje sustava za strojno prevođenje, a da oni pritom za to nisu plaćeni te često i ne znaju za što se sve nakon predaje upotrebljavaju ti prijevodi ili nemaju pravo izbora, čak i ako se osobno protive razvoju strojnoprevoditeljskih tehnologija (isto, 125).

Autor također izdvaja problematičnost izraza kojima se u istraživanjima opisuju uspjesi strojnih prijevoda, a koji potom pronadu put do drugih publikacija, marketinških materijala i medija: primjerice, u istraživanjima vrednovanja obično se gleda mogu li izlazni podaci postići „istovrijednost“ s konvencionalnim prijevodima, a ako izlazni podaci u vrednovanju postignu i bolje ocjene od referentnog ljudskog prijevoda, onda se govori o „nad-ljudskim“ izlaznim podacima (engl. *super-human output*), što u javnosti ostavlja dojam da je strojno prevođenje došlo do razine produkcije prijevoda savršene kvalitete i da ljudski prevoditelji stoga više nisu potrebni (isto, 128-129). I Thierry Poibeau argumentira da takvi izrazi, pogotovo kada se nađu u naslovima radova, javnost navode na pogrešan trag, a da se navodna istovrijednost strojnih i konvencionalnih prijevoda postiže samo na vrlo specifičnim vrstama tekstova, za vrlo ograničene jezične parove i u vrlo specifičnim uvjetima (2022: 6018). Također, istovrijednost ostaje navodna već zato što sustavi za strojno prevođenje funkcioniraju drukčije od ljudi. Dok se programi bore s pokušajima nošenja s koherencijom već unutar samog teksta koji prevode, ljudi su u stanju tekstove razumjeti na dvjema razinama odnosno uspostavljaju odnose i između elemenata unutar samog teksta (interna kohezija) i između elemenata teksta te stvarnog svijeta

(eksterna kohezija) (isto, 6020). Stoga je važno da istraživači osvijeste kako jezik snažno može utjecati na percepciju javnosti o strojnom prevođenju. S druge se strane može zaključiti da je i prevoditeljima u interesu da se upoznaju s istraživanjima i teorijskim okvirom strojnog prevođenja, kako se pri formiranju svojih mišljenja ne bi vodili medijskim napisima i nejasno interpretiranim podacima, već činjeničnim stanjem.

2.2 Vrednovanje strojnih prijevoda

Kako bi se kvaliteta nekog prijevoda, neovisno o tome je li ga izradio čovjek ili računalo, mogla poboljšati, ona se prvenstveno mora moći izmjeriti (Way 2018: 159). U postupcima vrednovanja otpočetak jednako sudjeluju i krajnji korisnici i razvojni programeri i istraživači, a kako će se kvaliteta procijeniti uvelike ovisi o prirodi i svrsi vrednovanja. Krajnji korisnici odabirući i vrednujući sustave za strojno prevođenje uvijek na umu imaju svoje svakodnevne potrebe i izabrat će onaj sustav koji najbolje odgovara konkretnoj situaciji u kojoj ga žele uporabiti; razvojni programeri s druge strane vrednovanjem pokušavaju ustvrditi prednosti i slabosti pojedinog sustava u određenoj razvojnoj fazi; a istraživači se ponajprije trebaju usredotočiti na to kako bi se sustav mogao unaprijediti tako da u budućnosti može odgovoriti na što različitiije potrebe korisnika (Nomura 1993: 206). Te tri vrste vrednovanja u načelu se vrše odvojeno, ali interakcijom među vrednovateljima mogu se dobiti odgovori čak i na ona pitanja koja metodom nisu bila ispitivana jer ljudski um može zapaziti i neke neočekivane i nepredviđene značajke sustava koje će dalje koristiti onima koji rade na njihovu razvoju.

Metode vrednovanja kvalitete ugrubo se mogu podijeliti u dvije skupine: ovisno o tome **što** se proučava odnosno **kako** se sustavi vrednuju dijelimo ih na takozvane *black box* i *glass box* metode vrednovanja, a ovisno o tome **tko** vrednuje sustave za strojno prevođenje govorimo o ljudskim i o automatskim metodama vrednovanja strojnog prevođenja (v. iduće potpoglavlje). Te podjele proizlaze iz različitih interesnih skupina koje provode vrednovanja: u znanosti se pokušava pronaći način da se dokaže napredak u kvaliteti prijevoda, a komercijalni je interes u istraživanju i razvoju sustava za strojno prevođenje doveo do proučavanja koliko je truda potrebno uložiti u redakturu prijevoda te na koncu i do razvoja automatskih metrika (Castilho i sur. 2018: 11).

Glass box metode vrednovanja tako se nazivaju jer se kvaliteta sustava za strojno prevođenje procjenjuje ispitivanjem njegovih unutarnjih komponenti, dok se kod *black box*

vrednovanja u obzir uzima samo krajnji rezultat, odnosno prijevod koji sustav ponudi (Dorr, Snover i Madnani 2011: 745). Pogled kroz „staklenu kutiju“ u načelu je informativniji za programere i tvorce sustava, dok „crna kutija“ ne pruža podatke o tome zašto se sustav ponaša na određen način, ali itekako mnogo govori o njegovoj funkcionalnosti i uspješnosti.

Black box metode najviše se temelje na usporedbi izlaznih podataka sustava s referentnim ljudskim prijevodom. Kvaliteta se, naime, procjenjuje tako da ljudski vrednovatelji proučavaju određene značajke prijevoda, najčešće tečnost i točnost, a pritom uglavnom uspoređuju strojni prijevod s referentnim ljudskim prijevodom (isto, 746). Kod odabira metode vrednovanja valja imati na umu da se sve metode „moraju prilagoditi svrsi za koju je pojedini prijevod napravljen, kao što i sustav mora biti razvijen za neku određenu svrhu“ (Vasconcellos 1993: 211). Gotovo je nerazumno očekivati da će za svaku svrhu za koju se vrednovanje provodi postojati ista, jednostavno provediva metoda vrednovanja. Naprotiv, uvijek će ih biti onoliko koliko ima i razloga da se vrednovanje uopće provede (isto).

2.2.1 Ljudske metode vrednovanja vs. automatske metrike

Da bi vrednovali kvalitetu, istraživačima je u cilju pronaći jednostavnu, a opet preciznu i valjanu metodu vrednovanja kvalitete strojnih, ali i ljudskih prijevoda. Neke od najboljih metoda i dalje su odveć mukotrpne (primjerice, ako se prijevod želi vrednovati testovima razumijevanja za čije je sastavljanje potrebno uložiti mnogo vremena i truda), a neke su pak odveć subjektivne ili jednostavno proizvoljne kada se određuju kriteriji kvalitete (Carroll 1996: 55). Potreba za brzim i jeftinim metodama velika je između ostalog i zato što se sustavi u fazi razvoja vrednuju više puta, pa je katkad već i nakon najmanje promjene u prototipu sustava poželjno provjeriti kako je ona utjecala na izlazne podatke (Nießen i sur. 2000: 1).

Iako je uvriježeno da ljudske metode vrednovanja imaju prednost nad automatskima, one su istovremeno u mnogim slučajevima i spore, skupe i nedosljedne (Way 2018: 165). Upravo zbog velikih troškova, subjektivnosti i sporosti razvijene su automatske metrike, a neke su od najpoznatijih: BLEU (*Bilingual Evaluation Understudy*), NIST (*National Institute of Standards and Technology*), METEOR (*Metric for Evaluation of Translation with Explicit Ordering*), WER (*Word Error Rate*), PER (*Position-independent Error Rate*), GTM (*General Text Matcher*), TER (*Translation Edit Rate*), CDER (*Cover Disjoint Error Rate*) i HTER (*Half Total Error Rate*) (Dorr, Snover i Madnani 2011: 749). Sve su ove metrike primjenjive u

brojnim područjima, no ključna im je zamjerka činjenica da su njihovi rezultati tek kvocijenti i brojke koje nam ne pružaju dovoljno smislenih informacija. Iz brojki se ne mogu iščitati nikakve specifične značajke pojedinih sustava i one zapravo ne odgovaraju na pitanja o njihovim prednostima i manama (Aikawa i Rarrick 2011: 332), već imaju smisla samo kada ih uspoređujemo jedne s drugima (Babych i Hartley 2008: 2134), a i tada samo u svrhu rangiranja sustava.

Automatske metrike funkcioniraju tako da se više strojnih prijevoda uspoređuje s referentnim ljudskim prijevodima te se sustavi zatim rangiraju od najboljeg do najlošijeg. Isprva se proučavala samo njihova leksička sličnost s referentnim prijevodima na površinskoj razini, a potom i jezične značajke teksta, poput onih sintaktičkih (vrste riječi, rečenični dijelovi i strukture) i dubinski semantičkih (dopuštajući sinonime, konotacije, parafraze itd.) (Han 2016: 7-8). Ipak, zamijećena su određena ograničenja povezana s referentnim prijevodima. Kada metrika WER računa broj nužnih koraka za uređivanje izlaznih podataka, najveću ulogu igra odabrani referentni prijevod (Nießen i sur. 2000: 1) jer će prijevod koji se uvelike razlikuje od referentnog biti ocijenjen lošom ocjenom iako iz teorije prevođenja znamo da ne postoji jedan apsolutni, savršeni prijevod, već i prijevod koji se u potpunosti razlikuje od odabranog referentnog može biti adekvatan. Automatske metrike „svako nepodudaranje s ljudskim prijevodom smatraju pogreškom, a ne mogućom i opravdanom alternativom“ (Babych i Hartley 2008: 2134). Osobito leksička sličnost zapravo „ne odražava sličnost u značenju“ (Han 2016: 7) jer se značenjska ekvivalencija ostvaruje na razini segmenta ili teksta, a ne pojedinih riječi. Također, katkad usporedba prijevoda pokaže da su npr. dva prijevoda djelomično kvalitetno prevedena, ali na različite načine, i automatske će metrike tada teško smisleno odrediti koji je sustav proizveo bolji prijevod (Dorr, Snover i Madnani 2011: 745).

Prednost ručnih metoda vrednovanja u odnosu na automatske metrike krije se u tome što se ljudi usputno mogu pozabaviti i složenijim jezičnim pojavama, kao i ustvrditi, primjerice, kakve se pogreške najčešće javljaju u nekom tekstu (Castilho i sur. 2018: 25). Automatske metrike ne razlikuju vrste pogrešaka, dok čovjek svjesno može zanemariti pogreške za koje procijeni da nisu presudne za ukupnu sliku o kvaliteti prijevoda. Neke pogreške mogu biti neznačajne, dok druge u potpunosti onemogućuju razumijevanje teksta (Dorr, Snover i Madnani 2011: 749). Tako, primjerice, pogreška u interpunkcijskim znakovima slabije utječe na razumljivost prijevoda od pogrešnog leksičkog odabira, osobito ako je pogrešno preveden neki ključni koncept. Katkad samo jedna pogrešna riječ može dovesti do potpunog nerazumijevanja, pa će

se ona smatrati i označiti kao ključna (Sudoh, Takahashi i Nakamura 2021: 47). Zato su apsolutne vrijednosti koje automatske metode nude kao rezultat svojih vrednovanja katkad jednostavno besmislene (Han 2016: 11), a ljudska zapažanja i dalje nezamjenjiva u području vrednovanja strojnog prevođenja. Upravo stoga ljudske se metode vrednovanja i unatoč subjektivnosti i proizvoljnom definiranju kriterija smatraju zlatnim standardom vrednovanja.

Još je jedno ograničenje automatskih metrika njihova (ne)osjetljivost kada im je zadatak rangirati nekoliko kvalitetnih prijevoda. Automatske metrike lakše rangiraju sustave kod kojih kvaliteta osjetno varira, ali ako procjenjuju više sustava dobre kvalitete, metrike će zakazati. S obzirom na to da kvaliteta izlaznih podataka svih sustava neprekidno raste te da sustavi sve više napreduju, može se pretpostaviti da će vrijednost automatskih metrika kakve danas poznajemo s vremenom opadati (Babych i Hartley 2008: 2133).

Zaključno se o automatskim metrikama može reći da su rano ispunile očekivanja brzine, podesivosti, replikabilnosti i niskih troškova, ali da je pred istraživačima i programerima još velik izazov u vidu razvoja metrika koje bi pružale smislene i ispravne rezultate kod kvalitetnijih prijevoda (Han 2016: 11). Upravo zbog toga i dalje su potrebna istraživanja ljudskih metoda vrednovanja koja jedina mogu držati korak s rapidnim napretkom strojnog prevođenja te koja mogu procjenjivati kvalitetu strojnih prijevoda za različite svrhe ili prema različitim kriterijima.

2.2.2 Kriteriji ljudskog vrednovanja strojnih prijevoda

Kvaliteta strojnih prijevoda vrednuje se prema unaprijed definiranim kriterijima. U ALPAC-ovu izvješću odabrana su dva kriterija: **razumljivost** i **točnost** prijevoda. Razumljivost se odnosi na to da prijevod mora biti jednako lako čitljiv i razumljiv kao što je rečenica čitljiva i razumljiva na izvornom jeziku⁴, a točnost se odnosi na to da se značenje teksta u prijevodu ne smije razlikovati od značenja teksta u izvorniku (Han 2016: 2-3). Kriteriji poput razumljivosti i točnosti, a poslije sve češće i **informativnosti**, kojim se provjerava je li kvaliteta strojnog prijevoda dovoljna da čitatelj iz njega može izvući tražene informacije (Han 2016: 3), ispituju se metodom u kojoj vrednovatelj pregledava pojedinačne segmente, a zatim ih ocjenjuje ocjenama s neke prethodno definirane ljestvice. Prosječna ocjena svih segmenata naposljetku pokazuje koliko su izlazni podaci razumljivi, točni ili informativni. Provjera informativnosti

⁴ Ovaj se kriterij na hrvatskom uvriježio pod nazivom „tečnost“ te se u ostatku rada rabi taj izraz.

može se provoditi i testovima s ponuđenim odgovorima, gdje uspješnost rješavanja sama po sebi svjedoči o tome sadrži li prijevod sve ključne informacije (isto). Važno je provjeriti može li čitatelj iz teksta izvući ključne informacije te u obzir uzeti težinu pogrešaka i namjenu prijevoda, kako bi se izbjeglo da vrednovatelji isključivo na temelju uočenih jezičnih pogrešaka strogo (pre)sude sustavu (Vasconcellos 1993: 212-213). Navedeni su kriteriji, točnost, razumljivost i informativnost, zapravo vrlo srodni i kroz prikaz se starijih radova vidi kako je postojala, a djelomično i danas još postoji, terminološka zbrka za pojmove koji su u optjecaju u teoriji vrednovanja strojnog prevođenja, što se osobito ističe u prijevodima na hrvatski. Tako su, primjerice, autori ALPAC-ova izvješća kriterij vrednovanja zapravo nazivali *vjernošću* (engl. *fidelity*) ili informativnošću (engl. *informativeness*), a u drugim se znanstvenim radovima na engleskom jeziku za isti koncept javljaju i drugi nazivi, poput adekvatnosti (engl. *adequacy*, v. Koehn 2009, Snover i sur. 2009, Castilho i sur. 2018 itd.) ili točnosti (engl. *accuracy*, v. Arnold i sur. 1994, Upute za uporabu sučelja MQM i dr.). Kod drugih se pak autora informativnost i adekvatnost razlikuju kao zasebni kriteriji. U ovom se istraživanju pokušava izbjeći ta terminološka nedosljednost te se rabe isključivo termini „točnost“ i „tečnost“, u skladu s dosadašnjom uvriježenom hrvatskom praksom. Značenja tih termina potanko su definirana u potpoglavlju 2.3.1.2 (Kriteriji kvalitete).

Dosad navedeni kriteriji vrednovanja usmjereni su na same prijevode, a ne na načine kako bi se mogla unaprijediti tehnologija koja je prijevode proizvela, a koju je jednako tako važno vrednovati. Margaret King, Andrei Popescu-Belis i Eduard Hovy (2003: 226) stoga donose pregled proširenog broja kriterija iz kasnijih istraživanja: **prikladnost** (odgovaraju li izlazni podaci određenom kontekstu u kojemu će se sustav rabiti?), **interoperabilnost** (funkcionira li sustav i na drugim softverskim ili hardverskim platformama?), **pouzdanost** („pada“ li sustav i koliko dugo se u tom slučaju ponovno pokreće?), **koristivost** (rabi li se sučelje sustava lako i intuitivno?), **učinkovitost** (koliko brzo prevodi sustav i kako ažurira promjene kada se prijevod revidira?), **mogućnost održavanja** (može li se sustav prilagođavati i mijenjati u skladu s potrebama korisnika?) te **prenosivost** (može li nova inačica sustava lako zamijeniti staru s obzirom na to da sustavi za strojno prevođenje nisu statični, već se brzo unapređuju i popravljaju?). Pritom je neke kriterije teže vrednovati od drugih – mnogo je teže, primjerice, zaključiti hoće li neki sustav odgovarati nekoj uporabi, djelomično i zato što metode za provjeru prikladnosti nisu osobito dobro razvijene (Arnold i sur. 1994: 157). Proizvođači sustava za strojno prevođenje imaju svoje metode vrednovanja navedenih kriterija, no u ovom je

lingvističkom radu fokus stavljen na *razumljivost* odnosno vrednovanje tekstualnog, a ne izvantekstualnih kriterija. U idućim se potpoglavljima izlaže pregled najčešćih ljudskih metoda vrednovanja strojnih prijevoda, čiji je polazišni kriterij obično točnost ili neka varijacija na točnost. Na svakom je istraživaču, jasno, da sam odabere prikladnu metodu za provjeru prikladnog kriterija, ovisno o namjeni provođenja vrednovanja strojnog prijevoda.

2.3 Metode vrednovanja strojnih prijevoda

Vrednovanje strojnih prijevoda kroz povijest se pokazalo „teškim, uznemirujućim i povremeno varljivim“ postupkom (White 2003: 211). Nikako nije jednostavno osmisliti pouzdane i učinkovite metode, osobito zato što za različite korisnike i uporabe moraju postojati i različite metode, a parametri poput brzine, kvalitete i cijene strojnog prevođenja ključni su u raspravama o tome treba li se i kako u prijevodnom procesu služiti automatiziranim komponentama (isto, 212). U sljedećim potpoglavljima donosi se prikaz prednosti i mana najučestalijih ljudskih metoda vrednovanja.

2.3.1 Metoda subjektivne procjene

Isti se koncepti na druge jezike mogu prevesti na mnoge različite načine te i se sami prevoditelji nađu u raspravama o tome kako bi se koji izvorni elementi najbolje izrazio na ciljnom jeziku. Razlog tomu bogata je varijabilnost jezika, kao i iznimna kreativnost koja je nužna za prevođenje (White 2003: 213). Metoda subjektivne procjene kvalitete prijevoda temelji se na činjenici da se subjektivnost kod vrednovanja ne može zaobići, ali se zato može iskoristiti. Subjektivne procjene kvalitete u načelu znače da vrednovatelji izlaznim podacima dodjeljuju ocjenu na nekoj unaprijed definiranoj bodovnoj ljestvici, koje pak mogu biti različitih stupnjeva robusnosti.

Za metodom subjektivne procjene posegnuli su već u ALPAC-ovu istraživanju u kojemu je sudjelovalo ukupno trideset i šestero ispitanika. Svim je ispitanicima engleski jezik bio materinski, a polovica ispitanika također je izvrsno vladala ruskim jezikom. Vrednovali su tri konvencionalna i tri strojna prijevoda na bodovnoj ljestvici od 0 do 9 po kriterijima točnosti i razumljivosti (ALPAC 1966: 81). Problem u takvim istraživanjima nastaje kada među različitim vrednovateljima ne postoji dovoljno visoka razina slaganja, odnosno kada

vrednovatelji zbog svojih stavova, iskustava, načina razmišljanja ili drugih razloga jednostavno imaju različito mišljenje o prijevodu koji se nalazi pred njima. Visoke razine neslaganja među vrednovateljima čine jedan od bitnijih izazova ove metode, zbog čega se, uostalom, u novije doba i odustalo od ljestvica ovako širokog raspona te prešlo na manje robusne ljestvice za koje se pokazalo da rezultiraju većim razinama slaganja među ispitanicima (Way 2018: 164).

Nakon prikupljanja podataka o ljudskim procjenama kvalitete, oni se moraju i smisleno analizirati, no ni ovdje se subjektivnost ne može zaobići. Ipak, kako su upravo ljudi ti koji se služe izlaznim podacima sustava za strojno prevođenje i kako samo oni sa svojim razumijevanjem svijeta mogu uočiti praktične pogreške u strojnim prijevodima, subjektivne se procjene i dalje naveliko rabe i ne postoji odgovarajuća zamjena za njih (Sanders i sur. 2011: 751).

Za ovakvo se istraživanje najprije moraju odabrati prikladni ispitanici. Istraživači se nalaze pred problemom bi li ispitanici trebali biti profesionalni prevoditelji, dvojezični laici ili samo jednojezične osobe te koliko bi ispitanike trebalo podučiti postupku vrednovanja strojnog prevođenja prije nego što im se povjeri taj zadatak. Znanstvenici znaju biti mišljenja da je za vrednovanje kvalitete potrebna određena teorijska pozadina, ali u jezičnoj industriji kvaliteta je pitanje stajališta i dogovora svih osoba koje sudjeluju u procesu isporuke prijevoda (poput poslodavca ili nadređenog, klijenta koji prijevod kupuje i, naravno, prevoditelja) te je stvar dogovora smatraju li oni da je ostvarena dostatna kvaliteta za neku određenu uporabu (Gouadec 2010: 270). Vrednovanje se u praksi znanstvenih istraživanja učestalo prepušta studentima ili laicima, a počешće i kolegama ili ljudima do kojih istraživači dođu preko mrežnih zajednica (Doherty 2017: 140).

Što se višejezičnosti tiče, glavno je ograničenje rada s jednojezičnim ispitanicima potreba za referentnim prijevodima. Njima se u tom slučaju pristupa kao savršenim prijevodima, premda neki dijelovi tih prijevoda mogu, primjerice, biti dvosmisleni, ili se jednostavno i u njima mogu potkrasti pogreške koje će se onda odraziti u vrednovanju (Sanders i sur. 2011: 756; Läubli i sur. 2020: 656). Ispitanici tada ujedno i nesvjesno vrednuju rad prevoditelja koji je napravio referentni prijevod te se njihov stav prema referentnom prijevodu može odraziti na procjenu kvalitete strojnog prijevoda (Ljubas 2020: 225).

2.3.1.1 Problem niskih razina slaganja kod subjektivnog ocjenjivanja prijevoda

Jedan od glavnih izazova metode subjektivne procjene je niska razina slaganja koja proizlazi iz činjenice da ne postoji jedan „idealni“ prijevod kojemu se teži, već i profesionalni prevoditelji za iste formulacije u izvornom jeziku mogu iznaći različite, jednako opravdane ekvivalente u ciljnom jeziku. Dosljednost ocjena odnosi se i na slaganje među različitim vrednovateljima (engl. *inter-annotator agreement*), ali i dosljednost vrednovatelja samih sa sobom (*intra-annotator agreement* prema Lommel, Popović i Burchardt 2014: 31), što je veoma teško ostvarivo. Načelno je poželjno imati što više vrednovatelja, ali razina slaganja tada zna opadati (Castilho i sur. 2018: 25). Jednostavno „nije moguće predvidjeti kakav stav i očekivanja ispitanici imaju kada pristupaju strojnome prijevodu“ (Simeon 2008: 141), a i pouzdanost svakog ispitanika može biti upitna jer, „ako je ispitanik skeptičan prema jezičnim tehnologijama ili ima averziju prema računalima uopće, vjerojatno će se to nepovoljno odraziti na njegovu ocjenu“ (isto). S druge strane, još pogubnije posljedice mogu imati pretjeran optimizam i preuveličavanje kvalitete izlaznih podataka jer vrednovanje „u javnu raspravu mora unijeti realan pogled na to što sustavi za strojno prevođenje mogu, a što ne mogu učiniti“, a također može pomoći predvidjeti „što bi bili u stanju učiniti u budućnosti“ (Yusof, Darus i Aziz 2017: 253).

Što se pak dosljednosti vrednovatelja samih sa sobom tiče, ne valja smetnuti s uma da je ocjenjivanje velikog broja rečenica uglavnom zamoran proces. Već i sami ispitanici zamjećuju da im je potrebno uložiti manje koncentracije za procjenu razumljivosti čitavih tekstova u odnosu na procjene razumljivosti svake pojedine rečenice nekog teksta, što se u ovakvim istraživanjima najčešće traži (Seljan i sur. 2015: 206)

U pokušajima da doskoče problemu nedosljednosti samih vrednovatelja i među vrednovateljima te odrede vjerodostojnost podataka dobivenih u istraživanjima, istraživači računaju koeficijent *Kappu*, Cronbachov koeficijent pouzdanosti ili neki drugi sličan koeficijent. Formula za računanje koeficijenta *Kappa* glasi: $K = \frac{P(A)-P(E)}{1-P(E)}$, gdje $P(A)$ označava koliko se puta vrednovatelji slažu u svom sudu, a $P(E)$ označava kolika je statistička vjerojatnost da se slučajno slažu⁵ (Callison-Burch i sur. 2008: 83). Iako koeficijent *Kappa* ne može spriječiti nesuglasje, ipak može ponukati istraživača da obrati pozornost na različitost

⁵ $P(E)$ uvijek ovisi o ljestvici kojom se vrednovatelji koriste, npr. $P(E) = \frac{1}{3}$ ako se koriste ljestvicom s 3 boda.

sudova vrednovatelja i da bude svjestan u kojoj su mjeri njegovi rezultati valjani. Istraživanja su pokazala da je koeficijent slaganja često niži kada vrednovatelji procjenjuju cjelovite rečenice, a da je viši kada se od vrednovatelja traži da rangiraju sustav ili procijene opću prihvatljivost prijevoda (isto, 85).

Cronbachov koeficijent pouzdanosti ima raspon od 0 do 1, gdje 1 podrazumijeva savršeno slaganje među vrednovateljima, koje se proporcionalno smanjuje kako se vrijednost približava ničtici. Sve ispod 0,5 smatra se neprihvatljivom razinom slaganja, između 0,5 i 0,6 slaganje je nisko, između 0,6 i 0,7 slaganje je prihvatljivo, između 0,7 i 0,9 slaganje je dobro, a sve iznad 0,9 smatra se odličnim (Seljan, Tucaković i Dunder 2015: 1094).

Posljednjih se godina u pogledu razina slaganja među ispitanicima propitkuje dotadašnja norma vrednovanja strojnih prijevoda na razini rečenice. Iako je u istraživanju koje su Sanja Seljan i sur. proveli na englesko-hrvatskom, njemačko-hrvatskom i rusko-hrvatskom jezičnom paru uočena statistički relevantna veća razina slaganja među ispitanicima koji su vrednovali raščlanjene segmente (2015: 206), Samuel Läubli i sur. proveli su empirijsko istraživanje na temelju kojega su zaključili da je vrednovanje na razini rečenice možda funkcioniralo dok je općenita kvaliteta izlaznih podataka bila vrlo niska, ali da je ispitanicima danas potrebno na uvid pružiti čitav tekst kako bi pažnju mogli usmjeriti i na nadrečenične nijanse značenja (2020: 659). Autori stoga preporučuju da se kao jedinica vrednovanja odabere cjelokupni dokument (primjerice puni novinski članak), a ne nasumično prikazane rečenice kao što je dosad bila ustaljena praksa, jer se ispitanicima treba pružiti mogućnost da uoče i pogreške koje se tiču ukupne kohezije i koherentnosti teksta, poput nedosljednih prijevoda imena istog proizvoda unutar istog teksta (isto, 668).

Isto predlaže i Melanie Mitchell, koja je zamijetila da odgovor na pitanje dostižu li sustavi za strojno prevođenje ljudsku razinu kvalitete uvelike ovisi o metodama vrednovanja i interpretiranju rezultata vrednovanja. Smatra da se, iako rezultati strojnog prevođenja u novije vrijeme jesu bolji, ne može govoriti o tome da dostižu ljudsku razinu i „uče semantičko značenje“ jer su rezultati subjektivnih procjena raščlanjenih segmenata „varljive“: dobru će ukupnu ocjenu dobiti i oni prijevodi kod koji su mnogi segmenti dobili najbolju ocjenu, ali su neki dobili najnižu (jer se uzima aritmetička sredina), ali i oni prijevodi kod kojih su svi segmenti dobili srednju ocjenu (Mitchell 2019: 191-193). U novije se vrijeme, stoga, sve češće mijenja način provođenja subjektivne metode vrednovanja.

2.3.1.2 Kriteriji kvalitete

Jezična kvaliteta teksta subjektivnom se metodom najčešće procjenjuje na dvije razine: točnosti i tečnosti (Rossi i Carré 2022: 57). Iako su, kao što je već spomenuto, kriteriju točnosti pridružena i neka druga imena (poput vjernosti, adekvatnosti ili informativnosti), on se uvijek odnosi na očuvanje izvornog značenja rečenice (Simeon 2008: 29). Točnost se ocjenjuje odgovorom na dva pitanja: „Prenose li izlazni podaci isto značenje kao i izvorna rečenica?“ i „Jesu li dijelovi značenja izostavljeni, dodani ili izmijenjeni?“ (Koehn 2009: 218). U to se također ubrajaju i neprevedene riječi koje se u izlaznim podacima pojavljuju na izvornom jeziku. Međutim, da bi se točnost nekog prijevoda uopće mogla ocijeniti, on čitatelju prvenstveno mora biti barem donekle razumljiv. Ako su izlazni podaci čitatelju posve nejasni, ostalim se kriterijima ne mogu dodjeljivati nikakve ocjene. Na razumljivost u jednakoj mjeri utječu i gramatička ispravnost, i odabir prijevodnih ekvivalenata i eventualne neprevedene riječi, a donositelj apsolutnog suda uvijek je čitatelj (Simeon 2008: 27-28). Kada vrednovatelj pak počne donositi ocjene za pojedine segmente, važno je da u obzir uzme količinu odstupanja od pravilnosti, a ne da ih samo procjenjuje kao „točne“ ili „pogrešne“ (isto).

Kriterij vrednovanja koji s točnošću čini jednu cjelinu jest tečnost. Tečnost se ocjenjuje odgovorom na pitanje „Jesu li izlazni podaci izraženi 'dobrim', tečnim jezikom?“, u što se ubrajaju i gramatička ispravnost te idiomatski odabir riječi (Koehn 2009: 218). U praksi se procjena tečnosti radi jednostavnosti uglavnom provodi s jednojezičnim ispitanicima: dostupniji su i mogu vrednovati izlazne podatke neovisno o izvornom jeziku. I u „Praktičnim smjernicama za uporabu MQM-a u znanstvenim istraživanjima o kvaliteti prijevoda“⁶ navodi se da je tečnost povezana s „jednojezičnim kvalitetama izvornog ili ciljnog jezika“ te da se procjeni tečnosti može pristupiti i posve zanemarujući činjenicu da je riječ o prijevodu (Burchardt i Lommel 2014: 14). U dokumentu se navode i specifične pogreške koje mogu utjecati na tečnost teksta, a one su: pravopisne pogreške i zatipci (između ostalog u to se ubraja i pisanje velikog te malog slova), tipografija (prekomjerna ili pogrešna uporaba pravopisnih i interpunkcijskih znakova) i gramatičke pogreške (sintaktičke pogreške, pogrešan oblik riječi, uporaba pogrešne vrste riječi, sročnost, pogrešno odabrano glagolsko vrijeme, način ili stanje,

⁶ „Practical guidelines for the use of MQM in scientific research on translation quality“, dostupno na: <http://www.qt21.eu/downloads/MQM-usage-guidelines.pdf> (pristup: 28. travnja 2020.). MQM odnosi se na metriku *Multidimensional Quality Metrics* (za detaljnije o metrici vidi potpoglavlje 2.3.3.3).

redosljed riječi u rečenici i neispravna uporaba funkcionalnih riječi poput prijedloga ili čestica).

Kriteriji točnosti i tečnosti ne ovise izravno jedan o drugom – prijevod može biti tečan čak i kada značenje izvornika nije vjerno preneseno, ali značenje može biti vjerno preneseno i kada tekst nije u potpunosti tečan. Još se jednom valja prisjetiti da se „točnost“ i „tečnost“ ovdje ne gledaju kao općejezični pojmovi kojima se i laički služimo kada opisujemo razne prijevode ili tekstove na izvornom jeziku, već su to jasno definirani i razgraničeni tehnički termini unutar discipline vrednovanja strojnog prevođenja. Tako se, primjerice, pogrešno odabrano glagolsko vrijeme kao pogreška ne pripisuje točnosti, već tečnosti prijevoda i očekuje da će dovesti do smanjivanja ocjene tečnosti, dok, s druge strane, izostavljanje neke riječi neće nužno utjecati na tečnost prijevoda, ali hoće na točnost (Snover i sur. 2009: 260). No oba kriterija moraju biti barem donekle zadovoljena kako bi izlazni podaci bili prihvatljivi. Stoga se kao ocjena kvalitete prijevoda uglavnom uzima aritmetička sredina ocjena dodijeljenih tim dvama kriterijima (isto, 259).

Norwati Yusof, Saadiyah Darus i Mohd Juzaidin Aziz (2017: 256-259) proveli su istraživanje u kojem su ispitanici procjenjivali koji se kriteriji moraju zadovoljiti da bi neki konvencionalni ili strojni prijevod bio prihvatljiv. Ispitanici su istaknuli da im je kod strojnog prijevoda najvažnije da „sadržaj prijevoda u cjelini bude lako razumljiv“⁷ (u čemu je sadržana točnost i tečnost prijevoda), a da im je najmanje važno „da se prijevod na ciljnom jeziku tečno čita“⁸. Činjenica da im je „tečno čitanje teksta“ manje važno pokazuje da su se čitatelji strojnih prijevoda spremni „probiti“ kroz tekst koji nije sasvim tečan ako će na kraju uspjeti pronaći traženu informaciju. Za konvencionalne prijevode ljudi pak intuitivno imaju veće kriterije, što se odražava i na njihovu vrednovanju. Kao najvažniju značajku konvencionalnih prijevoda ispitanici ističu „da prijevod mora sadržavati rečenice koje se prirodno čitaju u ciljnom jeziku“⁹. Nalazi ovog istraživanja odgovaraju postavkama Arnolda i sur. koji smatraju da se prava kvaliteta strojnog prijevoda krije u općoj razumljivosti: ako čitatelj ni uz trud ne može odgonetnuti značenje neke rečenice, prijevod mora dobiti lošu ocjenu, a tekstovi koji su vrednovateljima načelno jasni dobivaju bolje ocjene i unatoč uočenim jezičnim pogreškama (1994: 169).

⁷ „The content of a translation as a whole should be easy to understand.“ Koeficijent važnosti iznosi 3.26.

⁸ „A translation should read fluently in the target text language.“ Koeficijent važnosti iznosi 2.87.

⁹ „A translation should contain sentences that read naturally in the target language.“ Koeficijent važnosti iznosi 3.77.

2.3.1.3 Bodovne ljestvice za ocjenjivanje prijevoda

Jednostavan je postupak subjektivne procjene kvalitete: vrednovatelji na uvid dobivaju izlazne podatke nekog sustava za strojno prevođenje, a uz to najčešće i izvornik ili referentni ljudski prijevod, a zatim ih se zamoli da procijene kvalitetu tih podataka na nekoj ljestvici (Koehn 2009: 218-219). Zlatni standard vrednovanja nalaže da neka „bilingvalna osoba uspoređi izlazne podatke prevedene s izvornog jezika na ciljni i nekako oformi svoje mišljenje o kvaliteti prijevoda“ (Sanders i sur. 2011: 750). To nejasno „nekako“ pokušava se kvantificirati različitim bodovnim ljestvicama.

Ljestvica mora sadržavati najmanje dvije ocjene koje predstavljaju jedan pozitivan i jedan negativan sud, ali u praksi ipak najčešće sadrže četiri ili pet ocjena. Također je važno da su vrijednosti svih bodova nedvosmisleno naznačene kako bi ih vrednovatelji što sličnije tumačili, ali i da je jasno naznačeno koji se kriterij vrednuje na pojedinoj ljestvici. U skladu sa zaključkom da se prava kvaliteta krije u razumljivosti, Arnold i sur. (1994: 162, cit. prema Simeon 2008: 28) daju primjer ljestvice s četiri boda:

- 1 Rečenica je potpuno jasna i razumljiva. Ovjerena je i čita se kao običan tekst.
- 2 Rečenica je uglavnom jasna i razumljiva. Unatoč nekim netočnostima i nespretnim rješenjima, primatelj (gotovo) odmah može razumjeti što ona znači.
- 3 Opća misao je razumljiva tek nakon pomnijeg proučavanja. Rečenica sadrži gramatičke pogreške i/ili loš izbor riječi.
- 4 Rečenica je nerazumljiva. Iznalaženje njezinog značenja je beznadan zadatak; čak i ako se uzme u obzir kontekst, nagađanje se čini premalo pouzdanim.

Philipp Koehn (2009: 219) predlaže dvije ljestvice s ocjenama od 1 do 5 i ponešto jednostavnijim uputama. U ovom modelu vrednovatelji zasebno vrednuju točnost i tečnost (koja se u ovom slučaju odnosi na engleski jezik, ali se po istom načelu može primijeniti na bilo koji ciljni jezik):

Točnost		Tečnost	
5	puno značenje	5	tečan engleski
4	većina značenja	4	dobar engleski
3	mного značenja	3	engleski koji nije na materinskoj razini
2	malo značenja	2	engleski koji nije tečan
1	bez značenja	1	nerazumljiv engleski

Naposljetku, u istraživanju subjektivne procjene kvalitete strojnih prijevoda s engleskog na jezik urdu (Gupta, Joshi i Mathur 2013: 1521) uporabljena je ljestvica s vrijednostima od 0 do 4 s objašnjenjima koja su jednako primjenjiva za sve kriterije i za sve jezike:

- 0 Neprihvatljivo
- 1 Djelomično prihvatljivo
- 2 Prihvatljivo
- 3 Odlično
- 4 Savršeno

Mogućih je ljestvica mnogo i manje je važno koliko bodova ima, a više da su bodovi jasno razgraničeni i pojašnjeni. Premda će uvijek biti graničnih slučajeva kada ni sami ispitanici neće moći razlučiti što je za koju ocjenu, važno je ispitanicima pokušati sa što jednostavnijim i shvatljivijim sustavom ocjenjivanja olakšati odluku.

2.3.2 Metode provjere razumljivosti testovima razumijevanja

Korisnici strojnih prijevoda i sami se intuitivno zapitkuju je li im tekst koji čitaju razumljiv, a ako nije, brzo će odbaciti i njega, a možda i ideju strojnog prevođenja u cjelini. Prijašnja negativna iskustva sa strojnim prevođenjem te unaprijed utvrđeni stavovi mogu se odraziti na ocjene vrednovatelja kada ih se zamoli da procijene kvalitetu pojedinih segmenata. U nastojanjima da se subjektivnost neutralizira, uvedena je metoda vrednovanja testovima razumijevanja kojom se razumijevanje ispitanika pokušava objektivno izmjeriti i provjeriti (Sanders i sur. 2011: 752). Istraživanja upućuju na to da su testovi razumijevanja objektivni

pokazatelji kvalitete strojnih prijevoda neovisno o stavovima ispitanika, osobito kada im je vrijeme odgovaranja na pitanja ograničeno, primjerice, na jednu minutu (Fuji i sur. 2001: 105). Također znamo da razumijevanje uvelike može ovisiti i o tematici teksta: ponekad će čitatelji ovisno o svojim interesima, obrazovanju ili zanimanju moći biti u stanju razumjeti mnogo zahtjevnije tekstove, a neće mnogošto shvatiti iz jednostavnijih ili prosječno teških tekstova (Ghyselen 2017: 17).

2.3.2.1 Definiranje i ispitivanje pojma „razumljivosti“

Što zapravo znači razumjeti neki tekst? U Rječniku hrvatskog jezika „razumljivost“ je donekle tautološki definirana kao „svojstvo onoga što je razumljivo“, dok je prvo značenje glagola „razumjeti“ definirano kao „shvatiti smisao, bit“ (Šonje 2000: 1058). Razumljivost se u sjećanju osobe koja je pročitala neku rečenicu očituje na način da joj u glavi ostaje njezino značenje, ali ne nužno i površinska struktura (Sachs 1967: 437). Provjera istinskog razumijevanja mnogo je kompliciranija od prikupljanja subjektivnih dojmova o nekom tekstu. U nastavi stranih jezika profesori razumijevanje najčešće provjeravaju tako da učenicima nakon pročitanog teksta postavljaju pitanja. „Čitanje s razumijevanjem“ neizostavan je dio i većine standardiziranih ispita za provjeru razine znanja jezika¹⁰. Vrsta pitanja kojim se u tom dijelu provjere propituje razumijevanje varira ovisno o godinama i jezičnim razinama koje se ispituju.

Richard R. Day i Jeongyeon Park (2005: 4-5) predlažu taksonomiju koja im pomaže u osmišljavanju materijala za jezičnu poduku, a koja se sastoji od šest uočenih vrsta razumljivosti:

- (1) Doslovno razumijevanje (engl. *literal comprehension*) odnosi se na razumijevanje činjenica iz teksta na najpovršnijoj razini: činjenica poput datuma, vremena, lokacija i sl. Ova razina razumijevanja preduvjet je za sve ostale vrste.
- (2) Reorganizacija (engl. *reorganization*) temelji se na doslovnom razumijevanju, ali zahtijeva kombiniranje informacija iz različitih dijelova teksta.
- (3) Zaključivanje (engl. *inference*) zahtijeva kombinaciju doslovnog razumijevanja i izvantekstualnog znanja te intuicije. Zaključivanjem se dolazi i do odgovora koji nisu eksplicitno izrečeni u materijalu za čitanje, ali se u njemu podrazumijevaju.

¹⁰ Kao što su: za engleski jezik CAE (*Cambridge English: Advanced*), IELTS (*International English Language Testing System*) ili TOEFL (*Test of English as a Foreign Language*), za njemački jezik DSD (*Deutsches Sprachdiplom*), za španjolski jezik DELE (*Diplomas de Español como Lengua Extranjera*), za francuski jezik DELF (*Diplôme d'études en langue française*) ili DALF (*Diplôme approfondi de langue française*) itd.

- (4) Predviđanje (engl. *prediction*) također zahtijeva izvantekstualno poimanje o svijetu, a provjerava se tako što se čitatelju prekida čitanje na pola teksta i od njega traži da predvidi kako će se radnja nastaviti.
- (5) Vrednovanje (engl. *evaluation*) od čitatelja očekuje osvrt na pročitano, primjerice, očekuje se da čitatelj procijeni koliko mu je osobno neka informacija iz teksta korisna ili zanimljiva.
- (6) Osobno mišljenje (engl. *personal response*) jedini je tip razumijevanja kod kojega se odgovori na postavljena pitanja ne pronalaze u tekstu, već se temelje isključivo na čitateljevim osjećajima i mislima. No iako odgovori na pitanja osobnog mišljenja u načelu ne mogu biti pogrešni, razumijevanje teksta odražava se u odgovorima ako su argumentirani i povezani s činjenicama izrečenima u tekstu.

Autori uz ovu taksonomiju predstavljaju i pet vrsta pitanja s pomoću kojih se ove vrste razumijevanja mogu ispitati (isto, 6-8):

- (1) Jestna/niječna pitanja (engl. *yes/no questions*) služe za provjeru svih vrsta razumijevanja. Kao što ime nalaže, odgovor na njih uvijek je samo „da“ ili „ne“. Nedostatak ove vrste pitanja je što čitatelj ima izgleda od 50 % da pogodi točan odgovor.
- (2) Alternativna pitanja (engl. *alternative questions*) čitatelju nude dva ili više scenarija od kojih on mora odabrati jedan. Kao i kod jestnih/niječnih pitanja, nedostatak je taj što čitatelji imaju dobre izgleda da će pogoditi točan odgovor. Alternativnim se pitanjima mogu provjeriti sve vrste razumijevanja osim vrednovanja i osobnog mišljenja.
- (3) Pitanja točno/netočno (engl. *true or false*) konceptualno se naslanjaju na jestna/niječna pitanja. Iako i ovdje čitatelji imaju izgleda od 50 % da pogode točan odgovor, u nastavi jezika uz odgovore se najčešće moraju ponuditi i obrazloženja: ako premisa nije točna, zadatak je navesti razloge netočnosti, a ako je premisa točna, zadatak je objasniti zašto su ostali ponuđeni odgovori netočni. Priprema ovakvih pitanja zahtjevna je jer svi ponuđeni odgovori moraju zvučati vjerojatno i moraju se poigravati s nekim teže razumljivim aspektima teksta ne bi li čitatelja naveli na pogrešan trag i tako testirali njegovo stvarno razumijevanje teksta.
- (4) Pitanja koja se tvore s upitnim riječima (engl. *wh-questions*) odnose se na upitne riječi: gdje, što, kada, tko, kako i zašto. Čitateljima su osobito korisna pitanja s „kako“ i „zašto“ jer se odgovaranjem na njih odmiču od doslovnog značenja teksta i počinju ga dublje

razumijevati, ali se istovremeno smatraju i najtežima. Odgovor na ta pitanja uvijek zahtijeva izražavanje neke uzročno-posljedične veze, koja u tekstu može biti izražena na mnogo različitih načina, dok se odgovor na pitanje „tko“ lako već može raspoznati i na površinskoj razini ako je riječ o nekom imenu.

- (5) Pitanja višestrukog izbora (engl. *multiple-choice*) uvijek nude moguće odgovore među kojima čitatelj mora odabrati točan. Najčešće je točan samo jedan odgovor, ali ne uvijek. Ovakva su pitanja osobito prikladna za rad s govornicima koji su i dalje na nižim razinama poznavanja jezika.

Valja upamtiti da ovdje opisana pitanja, kao i vrste razumijevanja, nikako ne predstavljaju potpunu taksonomiju; autori, primjerice, na svoj popis nisu uvrstili zadatke u kojima se na prazne crte treba upisati odgovarajuća riječ (engl. *cloze test*) iako i njih rado rabe u nastavi, ali smatraju da su takvi zadaci prikladniji za „ocjenjivanje, a ne razumijevanje tipova razumijevanja“ koje predstavljaju u svom radu (isto, 6). No činjenica da je razumijevanje dobro proučen i poznat pojam iz prakse jezične poduke uvelike pomaže i u istraživanjima kvalitete strojnih prijevoda.

2.3.2.2 Tehnike ispitivanja razumljivosti u istraživanjima kvalitete strojnih prijevoda

Jednostavno rečeno, strojni je prijevod dobar onda kada su čitatelji teksta na ciljnom jeziku u stanju točno odgovoriti na pitanja o razumijevanju teksta, a loš kada to nisu u stanju učiniti (Scarton i Specia 2016: 3652). Pritom je vrlo važno kod pripreme testova razumijevanja voditi računa o tome koliko su pitanja koja će se nakon čitanja strojnog prijevoda postaviti ispitanicima zahtjevna i kompleksna jer strojni prijevodi koje na taj način vrednuju nisu nužno potpuno razumljivi i pitanja na vidjelo moraju izvući nijanse. Također pri pripremi testa treba pripaziti da uspješnost ispitanika ne ovisi o obrazovanju ili predznanju ispitanika o određenoj temi, već da odgovori proizlaze isključivo iz teksta. Sofia Bremin i sur. (2010: 48) provode istraživanje u kojemu ispitanici odgovaraju na pitanja višestrukog izbora, uz dodatak da na ljestvici od 1 do 8 izraze i koliko su sigurni u svoj odgovor kako bi se umanjila mogućnost pogađanja točnog odgovora. Slično od ispitanika traže Castilho i Ana Guerberof Arenas: nakon što odgovore na sva pitanja u testu razumijevanja, ispitanici se na ljestvici od 1 do 3¹¹

¹¹ 1 – ne slažem se, 2 – niti se slažem niti se ne slažem, 3 – slažem se

izjašnjavaju koliko se slažu sa sljedećim izjavama: (1) Tema teksta bila je lako razumljiva, (2) Jezik je bio lako razumljiv, (3) Pitanja su bila lako razumljiva, (4) Na pitanja sam odgovarao/odgovarala s velikom sigurnošću (2018: 82). Time su autorice u istraživanje uvele i element zadovoljstva ispitanika tekstom koji je ispred njih, umjesto da samo pobroje pitanja na koja su ispitanici ponudili točne odgovore.

Naime, već je u istraživanju Masarua Fujija iz 1999. godine dokumentirana diskrepancija između stavova ispitanika o kvaliteti prijevoda i broja točnih odgovora koje stvarno ponude na testu. U njegovu su istraživanju mnogi ispitanici izlazne podatke procijenili „nerazumljivima“ i „neprirodno napisanima“, ali su na pitanja ipak ponudili većinom točne odgovore, odnosno očito je da su ipak iz izlaznih podataka ipak iščitali sve ključne informacije (Fuji 1999: 289). I Ivana Simeon (2008: 126-127) u svom istraživanju bilježi veći broj ispitanika koja na sva pitanja u testu razumijevanja odgovaraju točno, ali prijevod ipak ocjenjuju nerazumljivim ili slabo razumljivim. Pritom valja imati na umu da ponekad ni izvornici nisu uvijek lako razumljivi i da ispitanici nisu upoznati s materijom, pa to može utjecati i na njihovu procjenu razumljivosti strojnoga prijevoda, odnosno on im se može činiti nejasnijim nego što zapravo jest (Dorr, Snover i Madnani 2011: 745).

Ispitanicima se mogu postaviti i pitanja otvorenog tipa kod kojih dužina željenog odgovora varira od nekoliko riječi do nekoliko rečenica (Meurers i sur. 2011: 1). Otegotna okolnost takvih testova jest ta što se zatim mora procijeniti i jesu li odgovori na pitanja otvorenog tipa smisleni i zadovoljavajući, što nije uvijek jednostavno. Odgovore je moguće bodovati na nekoliko načina, a i njihova točnost uvijek se procjenjuje u odnosu na unaprijed određeni „zlatni standard“ odgovora i nije apsolutna. Carolina Scarton i Lucia Specia definirale su sljedeći sustav bodovanja: ispitaniku se za točan odgovor (a najčešće je tu riječ o kopiranju dijelova teksta ili parafraziranju onoga što je u tekstu izrečeno) dodjeljuje jedan bod; za odgovor koji je načelno točan, ali sadržava i neki dodatni koncept koji nije točan dodjeljuje se 0,75 bodova; za odgovor koji nije cjelovit, već mu nedostaje neki važan koncept dodjeljuje se 0,5 bodova; za odgovor koji sadržava neki netočan koncept, a nedostaje mu neki drugi koncept koji bi upotpunio odgovor dodjeljuje se 0,25 bodova; i za potpuno netočan odgovor i odgovor koji ne proizlazi iz teksta dodjeljuje se nula bodova (2016: 3656).

Tehnika koja je na engleskome jeziku poznata kao *cloze procedure* provodi se tako da se određene riječi zacrne u tekstu, a onda se ispitanike zamoli da pogode koja riječ nedostaje. Kao metoda vrednovanja korištena je još 1960-ih i 1970-ih, a među njezine prednosti ubrajale su se

jednostavna provedivost i objektivni rezultati (Somers i Wild 2000: 1). Usuprot tomu, Susanne Göpferich (2009: 32-33) tvrdi da se ovom tehnikom mjere samo neki aspekti razumijevanja poput vjerojatnosti da će se neke riječi naći na nekom mjestu ili pak samo površnog razumijevanja cjelokupnog teksta. Dva je parametra važno uskladiti i dogovoriti prije provedbe ovakvog istraživanja: učestalost zacrnjivanja riječi u tekstu i kakvi se sve odgovori mogu prihvatiti kao točni. Učestalost zacrnjivanja riječi utječe i na ukupan broj riječi koje će u tekstu biti zacrnjene, a pitanje je trebaju li se riječi zacrniti nasumično ili u pravilnom razmaku. Harold Somers i Elizabeth Wild navode da je uobičajeno zacrniti svaku petu riječ, ali i da se može nasumično zacrniti primjerice 20 % riječi u tekstu (2000: 2).

Kao točan odgovor moguće je prihvatiti samo onu riječ koja je u tekstu i navedena, ali i sve druge odgovore koji gramatički i semantički odgovaraju na određeno mjesto u tekstu (primjerice, sinonime tražene riječi). Osobito je nezgodno kada se riječi u tekstu zacrnjuju u pravilnom razmaku, pa se dogodi da se zacrne vlastita imena, datumi ili druge činjenice koje je gotovo nemoguće pogoditi, neovisno o stvarnoj razumljivosti teksta. Međutim, zbog objektivnosti vrednovanja odgovora ispitanika autori su se u svom istraživanju odlučili za bodovanje samo onih odgovora koji su jednaki riječima u tekstu. Također su se odlučili zacrniti svaku petu riječ, no kako je u tekstu zbog toga ukupno trebalo ispuniti tristo praznina, ispitanici su se žalili na zamornost i zahtjevnost testa te su brojne crte ili ostavili prazne ili u njih upisivali besmislice. Toj zamornosti pridonijela je (ne)kvaliteta strojnog prijevoda u odnosu na kvalitetne tekstove na kojima se provodi ova tehnika kada se ispituje jezična razina učenika. Stoga su autori naposljetku došli do zaključka da je kod strojnih prijevoda bolje zacrniti tek svaku desetu riječ, odnosno da su u korpusu, koji je brojao petsto riječi, trebali zacrniti ukupno 10 % riječi (isto, 9).

Ekaterina Ageeva i sur. u svom su se istraživanju poslužili sličnom metodom. Umjesto upisivanja riječi na zacrnjena mjesta u strojnom prijevodu, istraživači su od ispitanika tražili da najprije pročitaju strojni prijevod, a onda popune praznine u referentnom, ljudskom prijevodu. Samo je jedna polovica ispitanika tijekom istraživanja na raspolaganju imala strojni prijevod (2015: 137). Nisu, dakle, vrednovali kvalitetu izlaznih podataka, već su se ovako usredotočili na to koliko je strojni prijevod koristan ispitanicima u razumijevanju izvornog teksta. Njihova hipoteza glasi da je „čitateljevo razumijevanje danog teksta u neposrednom odnosu s brojem riječi koje je sposoban rekonstruirati u izvorniku“ (isto, 138), a počiva na ideji asimilacije informacija (v. potpoglavlje 2.1.3) odnosno tome da strojni prijevodi prvenstveno služe u

informativne svrhe, pa se stoga može prihvatiti i slabija jezična kvaliteta. I u ovom se istraživanju kao najveći izazov pokazao odabir riječi koje bi se izostavile iz teksta jer je ključno odabrati one riječi koje prenose glavnu značenja cijele rečenice, što je prilično dugotrajan proces. Kod zacrnjivanja riječi dobro je u obzir uzeti spoznaju da glagoli i pridjevi znatno više utječu na razumijevanje od ostalih vrsta riječi (Maney i sur. 2012: 8).

Razumljivost strojnog prijevoda kategorija je koju se ne može odrediti pukim zbrajanjem pogrešaka te je stoga važno provoditi testove razumijevanja i unatoč tome što je njihova provedba skuplja od nekih drugih metoda vrednovanja (Jones i sur. 2007: 77). U istraživanju koje provode Douglas Jones i sur. ispitanici su odgovarali na test razumljivosti i strojnih i ljudskih prijevoda, a istraživači su točne odgovore vrednovali 1 bodom, djelomično točne odgovore s 0,5 bodova i netočne odgovore s 0 bodova. Unatoč velikom broju pogrešaka, strojni su prijevodi ispitanicima bili razumljivi u visokih 74 % slučajeva jer su glavni pojmovi bili točno prevedeni i to je ispitanicima omogućilo da točno odgovore na pitanja. Takvi se prijevodi nazivaju robusnim prijevodima (isto, 80). Za usporedbu, ispitanicima su ljudski prijevodi bili razumljivi u 95 % slučajeva, odnosno i u njima su zamijećene određene poteškoće. U brojnim stvarnim životnim situacijama dovoljni su i robusni prijevodi u kojima su osnovne informacije vjerno prenesene. Semantička se točnost, dakle, pretpostavlja gramatičkoj.

Da strojni prijevodi mogu biti korisni u svakodnevnom životu, kao i da je metoda testova razumijevanja prikladna za donošenje argumentiranih stavova o kvaliteti prijevoda, prikazano je u istraživanju koje su 2011. proveli Jan Berka, Martin Černý i Ondřej Bojar: ispitanici su čitali strojne prijevode praktičnih uputa i opisa životnih situacija (npr. upute dolaska do neke destinacije, aktualne vijesti, informacije o održavanju sastanka, softverske upute), a zatim odgovarali na preinačena jestno-niječna pitanja gdje su zapravo imali šest ponuđenih odgovora: (1) da, (2) vjerojatno da, (3) ne, (4) vjerojatno ne, (5) ne može se zaključiti na temelju teksta, (6) ne razumijem pitanje (2011: 79). Točni su odgovori zabilježeni u 80 % slučajeva. Također je na dva načina provedena provjera dosljednosti pojedinih vrednovatelja (*intra-annotator agreement*): u setovima pitanja postavljenima dijelu vrednovatelja pitanja su se ponavljala, i odgovori su pokazali prosječno slaganje od 92 %, a jedan je od vrednovatelja poslije dva mjeseca ponovno odgovorio na sva pitanja i pokazao razinu slaganja sa samim sobom u visini od 78,9 %, čime su podaci procijenjeni vjerodostojnima (2011: 81).

U istraživanju razumljivosti s neuronskim sustavom za strojno prevođenje (Macken i Ghyselen 2018) ispitano je razumijevanje jednog konvencionalnog prijevoda i dvaju

neuređenih strojnih prijevoda. Ispitanici su ocjenjivali razumljivost cjelokupnog teksta, ali su također označili i koji su im dijelovi teksta bili nejasni, a tu su se pokazale velike individualne razlike: neki su odlomci doista bili pogrešno prevedeni, drugi nisu bili izraženi posve idiomatski, dok se neke od tih poteškoća zapravo mogu povezati s izvornikom (2018: 123). Autori također pretpostavljaju da se razumijevanje značenja strojnog prijevoda u odnosu na konvencionalni razlikuje utoliko što čitatelj strojnog prijevoda mora nagađati dijelove značenja jer se u njima učestalo javljaju gramatički netočne rečenice (isto, 125).

2.3.2.3 Priprema testova razumijevanja

Istraživanja prikazana u prethodnom potpoglavlju pokazuju da testovi razumijevanja daju objektivnije podatke i da istraživači toj metodi pribjegavaju i unatoč tome što je teško pripremiti kvalitetne testove razumijevanja. No pripremu uvelike olakšavaju gotovi korpusi pitanja u okviru standardiziranih testova za ispitivanje razine stranih jezika. Većina instituta koji takve testove provode na svojim mrežnim stranicama na raspolaganje javnosti stavljaju svoje testove, pa se i istraživači njima mogu poslužiti (kao što su se, primjerice, Castilho i Guerberof Arenas u opisanom istraživanju iz 2018. poslužile materijalima testova IELTS-a). Takve testove sastavljaju profesori jezika i jezični stručnjaci koji se koriste stručnim znanjima i vještinama kako bi osmislili pitanja koja ciljano provjeravaju razumijevanje. Teme koje biraju uvijek su prikladne i izražene općim jezikom jer se od čitatelja nikada ne očekuju stručna znanja kako bi mogli razumjeti tekst. Stoga su takvi ispiti svakako prikladni i za istraživanja kvalitete strojnih prijevoda.

Postoje i drugi javno dostupni korpusi kojima se istraživači mogu poslužiti. Za rad s njemačkim jezikom javno je dostupan korpus CREG (*Corpus of reading comprehension exercises in German*). CREG sadrži više od sto izvornih dokumenata raznovrsne tematike s pripadajućim testovima razumijevanja koji najčešće sadržavaju pitanja otvorenog tipa. U načelu je namijenjen procjenjivanju razine znanja osoba koje njemački uče kao strani jezik, no njime se služi i sve više istraživača vrednovanja strojnog prevođenja. Uz pitanja je također uvijek naveden i odgovor koji predstavlja „zlatni standard“, ali koji može biti nadopunjen i drugim mogućim odgovorima. No i takvi se cjelovito preuzeti korpusi smiju prilagoditi potrebama istraživanja. Autorice Scarton i Specia, koje su se u opisanom istraživanju također služile korpusom CREG, prilagodile su tekstove tako da su jedan dokument podijelile u više testnih

materijala kako bi vrednovanje bilo manje zamorno, a k tome su izostavile dijelove tekstova koji su nepotrebno produljivali tekst, a nisu bili presudni za odgovaranje na pitanja i njihovo izostavljanje nije narušilo koherenciju teksta (2016: 3653). Pripremajući istraživanje ne smije se smetnuti s uma da se vrednovatelji lako umore ili zasite, te da će podaci biti pouzdaniji ako ispitivanje kvalitete ne traje predugo.

2.3.3 Metoda analize pogrešaka

Pogreška se u metodi analize pogrešaka u strojnim prijevodima definira kao „svaka poteškoća na koju čitatelj naiđe u prijevodu, a koja ili nije u skladu s izvornim tekstom ili se smatra netočnom u ciljnom jeziku“ (Burchardt i Lommel 2014: 12). Isprva se može činiti logičnim da se pogreške lako uoče i klasificiraju u određene kategorije, što je cilj metode analize pogrešaka. Ipak, anotatori se pogrešaka pritom suočavaju s brojnim izazovima, već i kod samog određivanja toga što je u određenom tekstu uopće pogreška, te je razina slaganja među njima i ovoj metodi u načelu prilično niska.

2.3.3.1 Izazovi provedbe analize pogrešaka

Razvojni programeri analizom pogrešaka pokušavaju doznati s čime njihovi sustavi za strojno prevođenje imaju najviše poteškoća (ili pak u čemu su već dovoljno dobri) kako bi znali koje njihove komponente još moraju modificirati ne bi li poboljšali rezultate. Vrednovanje im je ključno kako bi znali pomažu li njihovi naponi, i u kojoj mjeri, da sustav bolje prevodi (Arnold 1994: 172). S druge strane, prevoditelje i istraživače zanimaju i neke dodatne informacije o sustavu, npr. koji su problemi sustava za strojno prevođenje najizraženiji, koje su njegove konkretne prednosti, a koje mane, može li neka određena izmjena u sustavu poboljšati jedan određen aspekt prijevoda, pa makar to i ne utjecalo toliko na ukupnu ocjenu, je li možda najslabije ocijenjeni sustav u nečemu ipak bolji od najbolje ocijenjenog sustava, je li neke vrste pogrešaka teže ispraviti od drugih itd. (Popović 2018: 130). Najčešća je svrha analize pogrešaka izraditi profil pogrešaka u izlaznim podacima i rasporediti ih u unaprijed određene skupine pogrešaka (isto). No već je samo određivanje prikladne taksonomije pogrešaka prvi izazov u ovoj metodi jer se istraživač pita koliko detaljno valja odrediti pogreške da bi dobiveni podaci bili smisleni i informativni. Nekoliko čimbenika otežava izradu taksonomije: one „moraju

dobro odražavati i dobre i loše strane sustava za strojno prevođenje koje su važne za određeni zadatak i za određeni jezični par“, „opsežnije klasifikacije pružaju više informacija, ali je pogreške tako teže razlikovati“ i „vrste pogrešaka u klasifikaciji moraju pokrivati i lingvističke i prevoditeljske aspekte“ (isto, 132).

I uz jasno definirane taksonomije visoke se razine slaganja među anotatorima teško postižu jer se (1) anotatori ne slažu oko raspona teksta koji je zahvaćen pogreškom, (2) neke pogreške mogu objasniti na nekoliko različitih načina i nije jasno kako ih kategorizirati i jer se (3) anotatori katkad razlikuju u mišljenju oko toga je li nešto uopće pogreška i koliko je ozbiljna (Lommel, Popović i Burchardt 2014: 31). Vrlo čest primjer pogreške koja se može objasniti na nekoliko načina jest neslaganje glagola i imenice u broju: ako je imenica u jednini, a glagol u množini, ispitanik to može smatrati pogreškom u tečnosti, ali istovremeno i pogreškom u točnosti ako je npr. imenica izvorno u množini pogrešno prevedena imenicom u jednini. U radu s ljudima uvijek se mora uzeti u obzir da svatko od njih dolazi sa svojim jezičnim znanjem i shvaćanjem svijeta te da čak i analitički razrađeni okviri i detaljne upute kako odrediti pogreške ne mogu spriječiti neslaganja među anotatorima (isto, 36).

Otkada se istraživanja provode s neuronskim sustavima za strojno prevođenje, to je možda još i teže. Izlazni su podaci sustava za strojno prevođenje sve tečniji i gramatički ispravniji, pa i granice među pogreškama sve više blijede, odnosno klasifikacija pogrešaka postaje sve subjektivnija i sve više „stvar osobnog mišljenja anotatora“, a ne neke objektivne istine (Klubička, Toral i Sánchez-Cartagena 2018: 206).

Popović u svom radu donosi sažet pregled izazova analize pogrešaka (2018: 139-141), od kojih se kao najvažniji ističu:

- (1) profil anotatora: moguće je odabrati dvojezične ili jednojezične ispitanike, a u potonjem je slučaju potrebno osigurati referentni ljudski prijevod kako bi ispitanici mogli provesti analizu pogrešaka;
- (2) slaganje među anotatorima: istraživanja pokazuju da se slaganje među anotatorima u pitanjima raspona, težine i klasifikacije pogrešaka vrlo teško postiže, čak i s detaljnim uputama i obukom prije provedbe anotacije;
- (3) broj kategorija pogrešaka: detaljnije razrađene kategorije i potkategorije pogrešaka informativnije su, ali i otežavaju proces anotacije te dodatno narušavaju razinu slaganja među anotatorima;

- (4) definiranje kategorija pogrešaka: ne postoje univerzalne kategorije pogrešaka koje bi jednako dobro funkcionirale za sve jezične parove te za sve vrste tekstova;
- (5) ujednačavanje kategorija pogrešaka: ovaj je izazov povezan s definiranjem kategorija pogrešaka, ali se prvenstveno odnosi na činjenicu da se u prijevodima među određenim jezičnim parovima još javljaju segmenti koji toliko vrve pogreškama da ih zbog nerazumljivosti sve nije moguće ni anotirati – tada se pribjegava kategoriji „previše pogrešaka“ (ili „nerazumljivo“), no tada postoji dodatna opasnost da će anotatori segmente smjestiti u tu kategoriju čak i kada to načelno nije nužno. Kategorije je, dakle, nemoguće apsolutno ujednačiti jer svaki jezični par i svi izlazni podaci dolaze sa svojim posebitostima.

S ovim izazovima na umu, cilj je pripreme istraživanja stvoriti uvjete u kojima će anotatori moći postići optimalne rezultate kako bi bili vjerodostojniji za izvlačenje zaključaka (Bayerl i Paul 2011: 700). Pritom nikako ne valja smetnuti s uma da slaganje među anotatorima nije samo sebi svrha te da se može ostvariti i lažno slaganje, primjerice, ako se istraživanje provodi samo s nekoliko anotatora jer se oni mogu slagati i sasvim slučajno i nikako ne predstavljaju potpuni spektar mogućih mišljenja (isto, 713). Autori također dolaze do zaključka da u istraživanjima u jednakoj mjeri trebaju sudjelovati i profesionalni prevoditelji i jezičari te ljudi koji nemaju veze s prevođenjem jer se njihova mišljenja mogu jednostavno smatrati novim pogledima na jezične pojave i jednako su valjana kao i odluke stručnjaka (isto, 715-716).

2.3.3.2 Dobra praksa analize pogrešaka

Kako bi analize pogrešaka bile što vjerodostojnije, informativnije i izvedivije, znanstvenici su se usredotočili na pronalazak načina da se doskoči navedenim izazovima. Pritom valja imati na umu da rješenja nikada neće biti potpuna jer „sve ljudske metode vrednovanja strojno prevedenih tekstova imaju manu niske razine slaganje među vrednovateljima“, kao što su uostalom i sve metode ljudskog vrednovanja „dugotrajne i skupe“ (Stymne i Ahrenberg 2012: 1785), pa tako i analiza pogrešaka. Sara Stymne i Lars Ahrenberg u svom su istraživanju pokazali da se veće slaganje među anotatorima može postići na dva načina: služeći se jednostavnijim taksonomijama pogrešaka ili složenijim uputama za klasifikaciju određene pogreške (isto, 1789). Važno je, primjerice, uvjeriti se da anotatori razumiju da im je zadatak

označiti samo najmanji raspon teksta koji pogreška zahvaća – tako se eliminiraju barem ona neslaganja koja nastaju zbog nesporazuma, nepažnje ili nedovoljnog razumijevanja uputa.

Petra Saskia Bayerl i Karsten Ingmar Paul proveli su metaistraživanje o razini slaganja među anotatorima proučavajući prijašnja istraživanja analize pogrešaka. Na temelju analize (2011: 719) uočili su šest¹² čimbenika koji najviše utječu na razinu slaganja:

- (1) domena (uočene su varijacije u slaganju s obzirom na domenu tekstova),
- (2) broj anotatora (s većim brojem anotatora opada razina slaganja),
- (3) obuka (s intenzivnijom obukom prije istraživanja povećava se razina slaganja među anotatorima),
- (4) cilj (uočene su varijacije u slaganju s obzirom na cilj i provedbe anotacije),
- (5) stručnost anotatora u određenoj domeni (pokazalo se da homogene skupine stručnjaka u određenoj domeni i homogene skupine laika postižu veće razine slaganja od miješanih skupina stručnjaka i laika),
- (6) broj kategorija pogrešaka (s većim brojem kategorija smanjuje se razina slaganja među anotatorima).

Na temelju tog istraživanja, a s ciljem veće valjanosti podataka, autori predlažu uporabu taksonomija s manjim brojem kategorija pogrešaka, angažiranje anotatora koji su na jednakoj razini stručnosti u određenoj domeni te održavanje opsežne obuke za sve anotatore prije provedbe same analize (isto). Kod odabira anotatora važno je odrediti i hoće li se raditi s dvojezičnim ili jednojezičnim ispitanicima, što je upitno bilo i kod metode subjektivne procjene. Neka su istraživanja pokazala da jednojezični ispitanici vrednuju sporije, ali da je razina slaganja među njima veća nego kod dvojezičnih ispitanika kada su im samo dostupne informacije na ciljnom jeziku (Guzmán i sur. 2015: 464). Kako se dvojezični vrednovatelji smatraju rjeđima i skupljima, a i zato što je veća razina slaganja među vrednovateljima glavni cilj svakog istraživanja, rad s jednojezičnim ispitanicima zapravo je mnogo praktičniji (isto).

¹² Autori su zapravo proučavali osam čimbenika, no dva su u ovom pregledu izostavljena jer za njih nisu dobiveni relevantni podaci. Riječ je o čimbeniku jezičnih parova (nije zaključeno kako oni utječu na slaganje među anotatorima) te čimbeniku kalkulatorske metode (koja je povezana s analizom i predstavljanjem rezultata, a ne samom metodom klasifikacije pogrešaka).

2.3.3.3 Kategorije pogrešaka

U prilogu 11 ALPAC-ova izvješća prikazana je studija Arthura D. Littlea o kvaliteti materijala koji su strojno prevedeni s ruskog na engleski jezik. Zabilježene su ukupno 7573 pogreške koje su klasificirane u sljedeće kategorije (ALPAC 1966: 78):

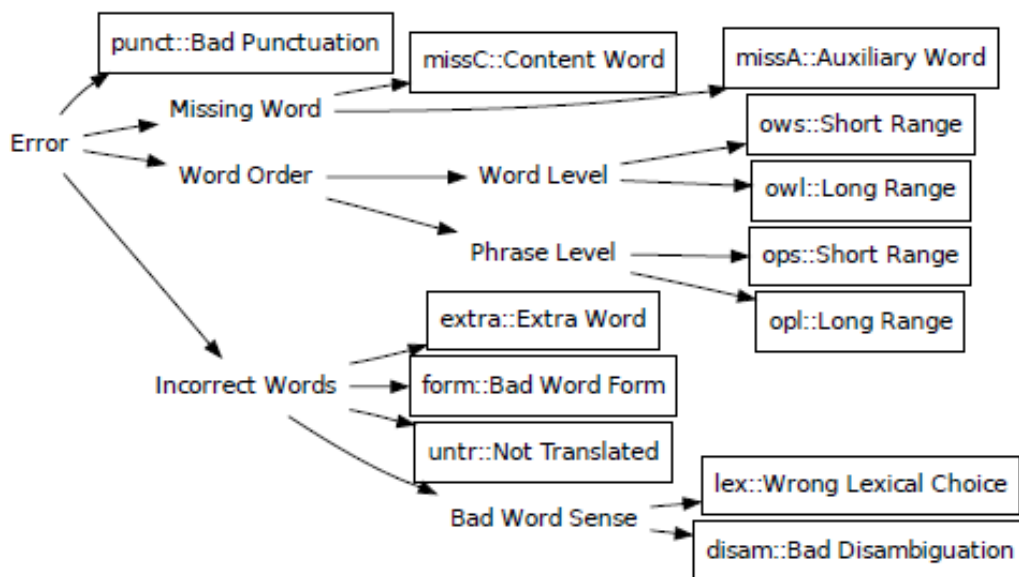
- (1) izostavljene riječi (s potkategorijama: „članovi“ i „ostalo“)
- (2) pogrešno odabrane riječi (s potkategorijama: „prijedlozi“, „glagolska vremena, stanja i nastavci“ i „ostalo“)
- (3) ostavljene ruske riječi
- (4) ostavljeno više izbora (podijeljeno ovisno o tome jesu li oba izbora pogrešna ili samo jedan)
- (5) suvišne riječi
- (6) simboli
- (7) neki dio rečenice uopće nije protumačen
- (8) redoslijed riječi

Ova se klasifikacija pogrešaka uvelike razlikuje od onih koje su danas uobičajene: kategorije nisu dobro razjašnjene, a pojedine su jako specifične, dok su druge preopćenite. Povrh toga ove kategorije ne obuhvaćaju sve moguće opcije pogrešaka.

Simeon (2008: 108) u svom istraživanju analizira pogreške u osam kategorija: „zadržavanje riječi iz izvornog jezika (neprevedene riječi), izostavljene riječi, suvišne riječi (neopravdano umetanje riječi ciljnog jezika), pogrešan leksički odabir (riječi koje se značenjski znatno razlikuju od riječi u izvornome tekstu), pravopisne pogreške, morfosintaktičke pogreške (pogrešni oblici, nesročnost subjekta i predikata, pogreške u članovima, pogrešno odabrane funkcionalne riječi i slično), stilističke pogreške (nespretno sročene sintagme, prijevodna rješenja koja su semantički bliska polaznoj riječi, ali mogu dovesti do nesporazuma ili zbuniti čitatelja) i pogrešan red riječi“. Ova je klasifikacija jasnija u odnosu na kategorije iz ALPAC-ova izvješća, to ne znači da su to sve moguće pogreške koje se mogu pojaviti u strojnim prijevodima, odnosno kategorije je uvijek potrebno prilagoditi. Na primjer, u istraživanju provedenom 2017. godine, izlazni su podaci prevedenog korpusa sa švedskog na hrvatski jezik tražili dodavanja dvije kategorija, premda je izvorno rabljena Simeonina klasifikacija kategorija za analizu pogrešaka (Ljubas 2017: 42-43): kategoriju pogrešaka u dosljednosti (jer je u izlaznim podacima zamijećeno da su isti stručni termini prevedeni na nekoliko različitih načina,

što je tekst činilo nejasnim) i kategoriju nejasnih rečenica (jer se u izlaznim podacima našao velik broj posve nekoherentnih rečenica koje su vrvjele isprepletenim pogreškama).

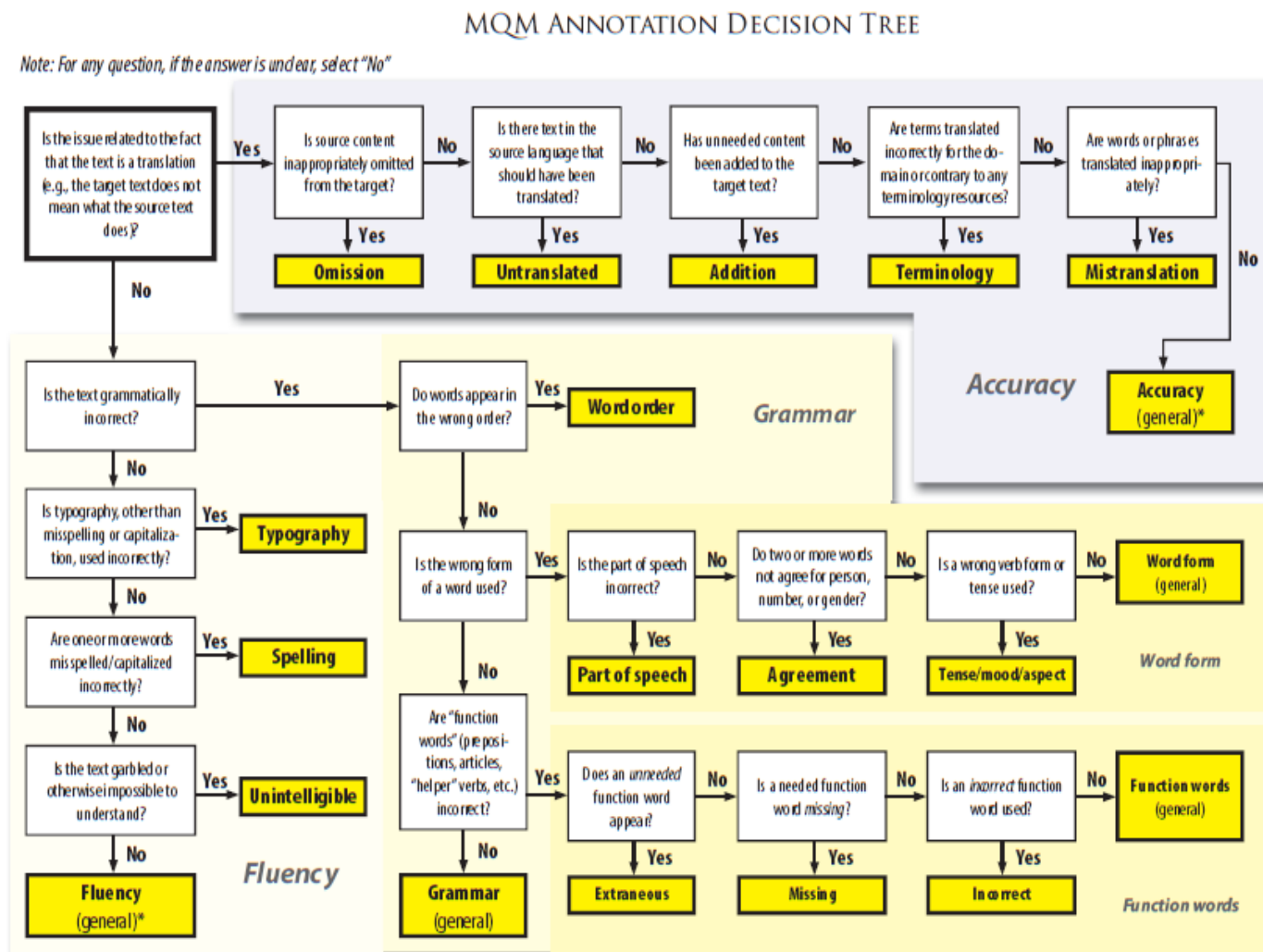
Slika 1 u nastavku prikazuje Bojarove kategorije za anotaciju pogrešaka (2011: 70). Kada anotator uoči pogrešku, smješta je u jednu od četiri glavne kategorije te odlučuje (1) je li riječ o pogreški u interpunkciji, (2) nedostaje li neka riječ, (3) je li redoslijed riječi pogrešan, ili su (4) odabrane riječi neispravne.



Slika 1. Klasifikacija pogrešaka u Bojarovoj studiji iz 2011. godine.

Nadalje se ulazi u potkategorije. Ako koja riječ nedostaje, anotator određuje je li riječ o punoznačnici. Kod redoslijeda riječi, anotator određuje je li riječ o pogreški na razini riječi ili dijela rečenice, i u sljedećem koraku dodaje informaciju o rasponu pogreške (kratki raspon znači da je riječ dovoljno pomaknuti za jedno sintaktičko mjesto u rečenici). Kategoriju neispravnih riječi Bojar dijeli na: umetnute riječi, neispravan oblik riječi, neprevedene riječi i na četvrtu, nešto nejasniju kategoriju, koja je predviđena za riječi koje se ne uklapaju u prijevod. Autor u posljednjem koraku nudi izbor između pogrešnog leksičkog odabira i pogrešno protumačene riječi. Razliku pojašnjava na način da je u potonjem slučaju razvidno da je sustav pogrešno shvatio izvorni tekst te odabrao riječ koja jest pogrešna, ali je u skladu s njegovim tumačenjem. Pogrešan je leksički odabir pak u potpunosti nemotiviran, odnosno nije jasno kako je sustav do njega došao (isto, 71-72). Te je dvije kategorije u praksi teže razlikovati od ostalih, što se dovodi u vezu s činjenicom da je upravo na tom mjestu razina slaganja među vrednovateljima bila najniža.

Na ovom tragu Aljoscha Burchardt i Arle Lommel u „Praktičnim smjernicama za uporabu MQM-a u znanstvenim istraživanjima o kvaliteti prijevoda“ iscrtali su vrste mentalnih mapa (engl. *decision trees*) s hijerarhijom pogrešaka (v. Sliku 2).



* Please describe any Fluency (general) or Accuracy (general) issues using the Notes feature.

Slika 2. Mentalna mapa za određivanje kategorije pogreške.

Kada anotator, primjerice, zamijeti neku pogrešku, prvo postavlja pitanje je li problem u nekom prevoditeljskom ili pak jezičnom aspektu. Točnost uvijek proizlazi iz suodnosa ciljnog teksta s izvornikom, dok se problemi tečnosti očituju samo u izlaznim podacima. Uzmimo primjer da anotator zabilježi prijevodnu pogrešku. Anotator tada prelazi na sljedeće pitanje: je li neki sadržaj izvornoga teksta izostavljen u prijevodu? Ako je odgovor potvrđan, anotator je odredio kategoriju pogreške. Ako je odgovor niječan, anotator nastavlja na sljedeće pitanje: nalazi li se u izvorniku još sadržaja koji nije preveden? Svaki potvrđan odgovor znači da je kategorija

pogreške uspješno određena, a svaki niječan vodi dalje: Jesu li u prijevod umetnute suvišne riječi? Jesu li termini pogrešno prevedeni u odnosu na terminološke konvencije? Jesu li riječi ili dijelovi rečenice neprikladno prevedeni? Ako je odgovor na sva pitanja niječan, anotator dopijeva na posljednje polje u kojemu mu je omogućeno upisati komentar i pobliže opisati uočenu poteškoću.

Druga je struktura po svojoj prirodi ponešto razgranatija, a ponuđene kategorije pogrešaka obuhvaćaju: ortografiju, pravopis, nerazumljivost, redosljed riječi, pogrešne oblike riječi (bilo da je riječ o pogrešnoj vrsti riječi, sročnosti ili obliku/vremenu, stanju ili vidu glagola) ili razne općenite probleme s gramatikom i tečnošću.

Ovakva sustavna pomoć pomaže u postizanju veće razine slaganja među anotatorima. Autori, međutim, ističu da se metrika ne mora uvijek i u svakom istraživanju razgranati u sve ove kategorije, već da anotatori mogu označavati samo one kategorije koje su relevantne za predmetna istraživačka pitanja, a idealno bi bilo kada bi se ponuđenih devetnaest kategorija svelo na njih pet ili šest (2014: 5). Osobito u novijim istraživanjima s neuronskim sustavima za strojno prevođenje pokazalo se da nije nužno imati, primjerice, toliko robusnu hijerarhiju tečnosti jer su problemi s tečnošću sve manji, ali je zato bolje uvrstiti više kategorija za kriterij točnosti (Klubička, Toral i Sánchez-Cartagena 2018: 213). S druge je strane zamislivo i da se Burchardtova i Lommelova hijerarhija pogrešaka proširi i na druge lingvističke razine poput stilistike ili pragmatike, gdje bi se anotatori pitali i o adekvatnosti prijevoda u određenom društvenom i kulturnom kontekstu. Ta je odluka uvijek ovisna o namjeni vrednovanja i kvaliteti izlaznih podataka, i važno je zapamtiti da su ova hijerarhija i sve slične njoj samo alati koji olakšavaju anotiranje pogrešaka, te da su unutar svakog istraživanja podložne promjenama.

2.3.3.4 Odnos zabilježenih pogrešaka i percepcije kvalitete

Istraživanja o tome koliki je utjecaj pojedinih vrsta pogrešaka na ljudsku percepciju kvalitete uglavnom su deskriptivnog tipa, no načelna je pretpostavka da pogreške koje se pojavljuju najčešće ne utječu nužno i najgore na percepciju kvalitete. U analizama pogrešaka pogreške se pokušavaju izdvojiti i tako se i broje, ali ne uzima se u obzir njihovo djelovanje u suodnosu (Federico i sur. 2014: 1643-1644). Utjecaj pogrešaka na percepciju kvalitete istražili su Marcello Federico i sur. u istraživanju s tri jezična para: englesko-kineskom, englesko-ruskom i englesko-arapskom. Jedna u radu neimenovana internacionalna organizacija ustupila je

autorima izvorne segmente na engleskom jeziku kao i tri strojna prijevoda segmenata na arapski, kineski i ruski jezik. Pokazalo se da u englesko-arapskom paru na percipiranu (ne)kvalitetu prijevoda najviše utječu leksičke pogreške, dok u englesko-kineskom i englesko-ruskom prijevodu percepciju kvalitete narušavaju neprevedene riječi, nadmašujući utjecaj i daleko češćih pogrešaka (isto, 1649).

Stoga se kod metode analize pogrešaka povremeno pogreškama pridružuje specifična postotna težina. Simeon među teže pogreške, tj. „one za koje se opravdano može pretpostaviti da znatno otežavaju razumijevanje teksta“ ubraja zadržavanje riječi iz izvornoga jezika, izostavljene riječi, suvišne riječi i pogrešan leksički odabir, dok među pogreške „koje u manjoj mjeri utječu na razumijevanje teksta“ ubraja pravopisne, morfosintaktičke i stilističke pogreške te pogrešan redoslijed riječi (2008: 123). Međutim, osnovni je problem kvantifikacije koeficijenata kojima bi se izrazila težina pogreške taj što nije jasno trebaju li koeficijenti odražavati utjecaj pogreške na razumljivost teksta ili pak broj koraka i zahtjevnost načina ispravljanja pogreške da bi tekst udovoljavao standardnim zahtjevima. Iako se zahtjevnost ispravljanja pogrešaka doista može mjeriti na razne načine, osobito je popularna Levenshteinova mjera udaljenosti kojom se računa najmanji broj znakova, umetaka, brisanja i zamjena koje se moraju poduzeti da bi se neka riječ, fraza ili rečenica promijenila u drugu (O'Brien 2022a: 116). Sharon O'Brien zorno to prikazuje na sljedećem engleskom primjeru: ako riječ *drink* treba ispraviti u *drunk*, potrebno je slovo „i“ zamijeniti slovom „u“; ali ako frazu „*He drinks*“ treba zamijeniti u „*He is drinking*“ Levenshteinova mjera udaljenosti je 6 jer je potrebno umetnuti slovo „i“, slovo „s“ i jedan razmak, zatim zamijeniti slovo „s“ slovom „i“, umetnuti slovo „n“ i umetnuti slovo „g“ (isto). No katkad, iako je neke pogreške relativno jednostavno ispraviti, one ipak mogu sa sobom povlačiti i druge pogreške, pa tako zahtjevi dodatno rastu (Simeon 2008: 30-31).

Arnold i sur. istaknuli su tri temeljna problema određivanja koeficijenta težina (1994: 165-166):

- (1) postupak analize pogrešaka već je dovoljno mukotrpan, a da bi anotatori analizirali koje pogreške konkretno utječu na slabo razumijevanje morali bi proći i mnogo dužu i težu obuku, te provesti daleko više vremena anotirajući;
- (2) kriteriji za odabir koeficijenta težine mogu se razlikovati i nije očito koje odabrati;

- (3) u prirodnom jeziku svaka riječ stoji u suodnosu s nekom drugom, pa je anotatorima katkad teško odrediti granice među pogreškama – što se najjasnije vidi u problemima sročnosti.

Svi su ti problemi već dugo poznati, ali za njih još nisu pronađena adekvatna rješenja. U današnje se vrijeme stoga sve češće pribjegava postupku redakture teksta koje se zapravo može smatrati implicitnom analizom pogrešaka (Popović 2018: 145).

2.3.3.5 Metoda redakture strojnog prijevoda

Osim u prevoditeljskoj praksi, postupak redakture strojnih prijevoda zaživio je i kao metoda vrednovanja kvalitete strojnih prijevoda koja je proizašla iz analize pogrešaka. Polazi od pretpostavke da su prijevodi koji zahtijevaju manje redakture zacijelo kvalitetniji, a manje su kvalitete oni koji zahtijevaju više redakture, no kao i kod klasične analize pogrešaka, najveći je izazov odrediti koje su pogreške bezazlene, a koje ključne za percepciju kvalitete (Dorr, Snover i Madnani 2011: 748-749). Zadatak je prevoditelja pokušati unijeti najmanji broj izmjena koji bi izlazne podatke pretvorili u prihvatljiv prijevod (Han 2016: 4), ali od njih se može tražiti da usput sastavljaju i popise učestalih pogrešaka koji se onda, zajedno s prijedlozima za poboljšanje takvih pogrešaka, prosljeđuju razvojnim programerima. Tako se praktična primjena spaja s istraživačkom vrijednošću podataka koje prevoditelji kroz svoj rad prikupljaju, ali taj je postupak kognitivno zahtjevan jer se sastoji od utvrđivanje pogrešaka, planiranje koraka za njihovo ispravljanje, ali i od tehničkih izazova, odnosno svih onih računalnih operacija koje se moraju izvršiti da se prijevod pročisti od pogrešaka (Klings 2001, cit. u Koponen 2012: 181). Neke se pogreške tek uoče nakon opsežnije analize, dok je druge lako uočiti, ali je potrebno poduzeti puno koraka kako bi se ispravile. Kognitivni naponi ove metode moraju se smanjiti kako bi se redakture mogle provoditi što brže (isto).

Istraživanja su pokazala da je morfološke pogreške najjednostavnije ispravljati, a potom slijede leksičke pogreške, naročito one koje se tiču stila i registra, zatim netočno prevedene riječi, umetnute pa izostavljene riječi i pogrešno prevedeni frazemi. Najteže je ispravljati pogreške koje utječu na sintaksu: pogrešnu ili izostavljenu interpunkciju, redosljed pojedinih riječi u rečenici i naposljetku redosljed riječi unutar rečenice ili skupova riječi (Koponen 2012: 182). Maarit Koponen od ispitanika istraživanja traži da rečenicama pridružuju ocjene

koje izražavaju koliko je zahtjevno bilo ispraviti pogreške u njima. Rezultati pokazuju da su najniže ocjene dobile rečenice s pogreškama u redosljedu riječi, odnosno da ispitanici percipiraju da je njih najteže ispraviti (isto, 188). Kao što je već spomenuto, ispitanici su kao lakše percipirali one rečenice u kojima su uglavnom ispravljali morfološke pogreške. Rezultati istraživanja također pokazuju da ispitanici dugačke rečenice percipiraju kao osobito teške za ispravljanje, iako je katkad broj ispravaka u njima razmjerno malen. Njihova su razmišljanja važna jer pokazuju da su kognitivni naponi za utvrđivanje pogrešaka u dugačkim rečenicama i planiranje kako ispraviti te pogreške mnogo veći nego što se može iščitati iz zbroja tehničkih koraka. To ujedno znači da bi se postupak redakture olakšao kada bi se tekst prikazivao u manjim segmentima (isto, 189).

Naposljetku, i redaktura je strojnih prijevoda vještina koja se može naučiti. Elizabeth Wagner još je 1985. godine sastavila upute za brzinsku redakturu strojnih prijevoda za potrebe prevoditeljske službe Europske komisije. Ti se prijevodi ne objavljuju, već se samo privremeno upotrebljavaju u radu Komisije. Pripremaju se u velikoj vremenskoj stisci i moraju se što brže dovesti na manje-više prihvatljivu razinu. Njezine upute glase (1985: 6):

- „Ako je moguće, služite se programima za obradu teksta i prijevod ispravljajte na računalu umjesto ručno.
- Nemojte ispravljati prijevod prije nego što, odlomak po odlomak, pročitate izvornik.
- Unosite samo najnužnije izmjene; ispravite besmislene i netočne riječi i izraze, a ako imate dovoljno vremena, onda i dvosmislene.
- Zadržite što više sirovog prijevoda i oduprite se kušnji da previše brišete i prepravljate.
- Dopustite slobodne prijevode i kolokvijalne izraze kada je jedina svrha teksta da prenese informaciju.
- Ne zadržavajte se predugo na nekom određenom problemu, označite ga i eventualno se poslije na njega vratite.
- Ne zamarajte se stilom prijevoda.
- Nemojte brzinski redigirati tekstove nepoznate tematike kako ne biste izgubili mnogo vremena rješavajući osnovne terminološke probleme.
- Nemojte dugo istraživati nešto što ne znate. Brzinski pretražite izvore terminologije, a ako problem ne riješite, samo ga označite upitnikom na marginama.

- Ne doživljavajte brzinsko redigiranje strojnih prijevoda preozbiljno i nemojte se laćati tog zadatka ako vas prijevodi niske kvalitete uzrujavaju.“

Ni danas se smjernice ne razlikuju drastićno, no TAUS (*The Translation Automation User Society*) razlikuje smjernice za laganiju ili površnju redakturu strojnih prijevoda (od koje se kao rezultat očekuje „dovoljno dobar“ prijevod) od uputa za punu redakturu (koja bi dovela do prijevoda ćija je kvaliteta usporediva s kvalitetom konvencionalnih prijevoda). Osnovne smjernice glase (Massardo i sur. 2016: 17):

- Usredotoći se na semantićki toćan prijevod.
- Pripazi da nijedna informacija nije slučajno dodana ili izostavljena.
- Uredi sav uvredljiv, neprikladan ili kulturološki neprihvatljiv sadržaj.
- Zadrži što je više moguće sirovih izlaznih podataka.
- Uvrsti osnovna pravila pravopisa.
- Nije potrebno unositi ispravke koje služe isključivo popravljanju stila.
- Nije potrebno restrukturirati rećenice samo da bi prirodnije tekle.

Dopunjene smjernice između ostalog još traže usmjeravanje pažnje i na gramatićku i sintaktićku ispravnost, interpunkciju, ispravno prevedenu terminologiju i ispravno formatiranje teksta (isto, 18). I standard za redakturu strojnih prijevoda „ISO 18857:2017“ takoder razlikuje osnovnu redakturu, definirajući je kao proces kojim se pokušava dobiti samo razumljiv tekst, koji ne pokušava parirati konvencionalnom prijevodu, od potpune redakture od koje se na koncu očekuje proizvod koji bi bio usporediv s prijevodom koji bi napravio ćovjek (ISO 2017: 2).

Osim što se prijevodi naknadno uređuju, moguće je tekst i unaprijed pripremiti za strojno prevođenje kako bi izlazni podaci bili što kvalitetniji (engl. *pre-editing*). Prirodni je jezik, naime, po svojoj naravi dvosmislen i toliko kompleksan da i računala i ljudi katkad imaju poteškoća s razumijevanjem određenih dijelova teksta, a tom se problemu može doskoćiti uporabom kontroliranog jezika, osobito kod tehnićke dokumentacije (Nyberg, Mitamura i Huijsen 2003: 245). Kontrolirani je jezik jasno odrećeno ogranićenje prirodnog jezika u smislu leksika, gramatike i stila kako bi tekstovi pisani kontroliranim jezikom postali lakše ćitljivi i razumljiviji, što pak u sljedećem koraku dovodi i do boljih konvencionalnih i boljih strojnih

prijevoda (isto). Eric Nyberg, Teruko Mitamura i Willem-Olaf Huijsen predlažu kako tekst valja napisati da bi bio pogodan za strojno prevođenje:

- Nemojte se služiti rečenicama dužima od dvadeset riječi.
- Nemojte se služiti pasivom.
- Nemojte slagati imeničke skupove duže od četiri imenice.
- Neka jedna rečenica sadrži samo jednu uputu.
- Vaše upute trebaju biti što konkretnije.
- Dugačke popise oblikujte s oznakama. (isto, 247).

Važnost kontroliranog jezika u olakšavanju strojnih prijevoda prepoznale su i neke tvrtke, među kojima se osobito ističe tvrtka Perkins Engines koja je pokušala pojednostaviti svoje publikacije na engleskom za neizvorne govornike, pa je još 1980. predstavila rječnik PACE (*Perkins Approved Clear English*) s dvije tisuće petsto riječi te nizom od deset pravila pisanja (Pym 1990: 85-86):

- Rečenice neka budu kratke.
- Izbacite zalihosne riječi.
- Rečenični dijelovi moraju biti logično posloženi.
- Ne mijenjajte konstrukcije usred rečenice.
- Obratite pažnju na logičnu uporabu riječi „i“ i „ili“.
- Izbjegavajte elipse.
- Ne izbacujte veznike ni odnosne riječi u službi veznika.
- Držite se rječnika PACE.
- Izbjegavajte nizanje imenica.
- Ne služite se riječima na „-ing“, osim ako su natuknice u rječniku PACE.

Ako bi se ovakve i slične upute dosljedno slijedile, pomoglo bi to sustavima za strojno prevođenje da eliminiraju pogreške te bi strojni prijevodi posljedično mogli biti mnogo uspješniji, zahtijevali bi sve manje redakture, a u najboljem slučaju – nimalo (Nyberg, Mitamura i Huijsen 2003: 254-255). To, naravno, ne znači da se svi tekstovi moraju pisati prateći ovakva stroga pravila, a ni da će kontrolirani jezik zamijeniti kreativni, prirodni jezik.

Ovakva primjena strojnog prevođenja najprimjerenija je za tehničku dokumentaciju i slične publikacije koje odišu izravnošću i jasnoćom.

3. Cilj i istraživačka pitanja

Budući da je glavni nedostatak svih ljudskih metoda vrednovanja niska razina slaganja među vrednovateljima, čime se u pitanje dovodi objektivnost rezultata, ishodišno je pitanje ovog rada kako se odabirom pravilnih metoda i promjenama parametra istraživanja može utjecati na to da dobiveni podaci o kvaliteti i razumljivosti prijevoda budu što vjerodostojniji i smisleniji, odnosno koji su optimalni uvjeti provedbe vrednovanja. Stoga će se na istom materijalu provesti istraživanje s različitim ljudskim metodama vrednovanja (metode subjektivne procjene kvalitete, testova razumijevanja i analize pogrešaka) da bi se usporedilo njihove prednosti i mane te proučilo kako se podaci dobiveni svakom pojedinom metodom razlikuju, s ciljem razvijanja bolje svijesti o tome kako odabrati i kako primijeniti koju metodu da bi se provelo valjano vrednovanje. Ovaj je cilj osobito važan u dobu rapidnih promjena u kvaliteti izlaznih podataka, kada se ne postižu svim metodama istovrijedni rezultati.

Ovim će se istraživanjem pokušati dati odgovor na sljedeća istraživačka pitanja:

- (IP1) Postiže li se veće slaganje među ispitanicima kada oni ocjenjuju prijevod u cjelini u odnosu na raščlanjene segmente ili obrnuto?*
- (IP2) Ovisi li ocjene pri procjeni razumljivosti teksta o prijašnjim iskustvima sa strojnim prevođenjem ili o stavovima o njemu?*
- (IP3) Jesu li ocjene pri procjeni razumljivosti teksta proporcionalne broju točnih odgovora na testu razumijevanja?*
- (IP4) Ovisi li ocjene pri procjeni razumljivosti pojedinih segmenata teksta o broju jezičnih pogrešaka u njima?*
- (IP5) Utječu li pogreške u leksičkoj točnosti bitnije na nerazumijevanje prijevoda od ostalih pogrešaka?*
- (IP6) Postiže li se smanjenjem broja jezičnih pogrešaka u izlaznim podacima veća dosljednost u rezultatima vrednovanja metodama subjektivne procjene i testa razumijevanja?*

Jedan od najvažnijih parametara metode subjektivne procjene raščlanjenost je segmenata pri vrednovanju. Procjenjivanje kvalitete velikog broja segmenata zahtjevan je i kognitivno naporan proces, a pad koncentracije kod vrednovatelja dovodi do niskih koeficijenata slaganja ispitanika samih sa sobom (v. potpoglavlje 2.3.1.1). Vrednovatelj koji zbog zamornosti zadatka

ne uspijeva zadržati pregled nad time primjenjuje li što ujednačenije kriterije pri vrednovanju svakog segmenta nedosljedan je i sam sa sobom, zbog čega nikako ne može ostvariti ni relevantnu razinu slaganja s drugim vrednovateljima. Prvo se istraživačko pitanje (IP1) naslanja na istraživanje u kojemu su ustvrđeni toliko niski koeficijenti slaganja među ispitanicima, čiji je zadatak bio ocijeniti opću prihvatljivost raščlanjenih segmenata strojnih prijevoda, da su autori za daljnja istraživanja ostavili otvoreno pitanje je li možda takvo vrednovanje toliko kompleksan zadatak da i nije moguće postići ikakvo slaganje (Callison-Burch i sur. 2008.). U recentnijim se istraživanjima sve češće preporučuje vrednovanje čitavih tekstova (na temelju rezultata svog empirijskog istraživanja promjenu u načinu vrednovanja osobito snažno preporučuju Läubli i sur. 2020, v. poglavlje 2.3.1.1), ali u kontekstu istraživanja strojnog prevođenja u kombinaciji s hrvatskim jezikom dosad još nije proučen utjecaj tog parametra na dosljednost.

Drugo istraživačko pitanje (IP2) odnosi se na međudjelovanje varijable unaprijed određenih stavova ispitanika o strojnom prevođenju i dodijeljenih ocjena prijevodu subjektivnom procjenom, a formirano je na istom tragu kao i treće istraživačko pitanje (IP3) u kojemu se iste te ocjene subjektivne procjene pak stavljaju u suodnos s brojem točnih odgovora na testu razumijevanja. Oba se istraživačka pitanja naslanjaju na rezultate Fujijeva istraživanja iz 1999. i Simeonina istraživanja iz 2008. (v. potpoglavljje 2.3.2.2) gdje su ispitanici prijevode u subjektivnoj procjeni ocijenili niskim ocjenama, a na testovima ipak pokazali visoke razine razumijevanja, što se objasnilo snažnim utjecajem unaprijed utvrđenih stavova ispitanika na rezultate vrednovanja. Kako je riječ o nešto starijim istraživanjima, postavlja se pitanje je li to i dalje tako, kao i ima li kakvih pomaka u percepciji, stavovima ili iskustvima ispitanika vezano uz strojno prevođenje. Pritom valja napomenuti da je, kao i kada negativni stavovi ispitanika prema strojnom prevođenju utječu na ocjene, jednako nepoželjno kada ispitanici prijevodima daju bolje ocjene nego što zaslužuju jer su osobno prema strojnom prevođenju pozitivno nastrojeni i voljeli bi ga pokazati u što boljem svjetlu. I za metodu subjektivne procjene i za metodu testa razumijevanja vrlo je važno da ispitanici odgovaraju iskreno i neopterećeno, ne razmišljajući o tome što se od njih očekuje, što bi bio ispravan odgovor i kakve će biti implikacije rezultata istraživanja.

Sljedeća se dva istraživačka pitanja (IP4 i IP5) odnose na analizu pogrešaka, IP4 na broj pogrešaka, a IP5 na utjecaj vrste pogrešaka. Analizom pogrešaka ne vrednuje se razumljivost sama po sebi, već se proučava jezična ispravnost teksta koja se potom stavlja u odnos s

razumljivošću. IP4 polazi od pretpostavke da velik broj pogrešaka ujedno znači da je tekst vrlo nerazumljiv, a manji broj pogrešaka da su izlazni podaci sustava za strojno prevođenje bolji, premda postoje i podaci koji govore da je katkad jedna jedina pogreška dovoljna da naruši čitavu razumljivost rečenice, primjerice s pogrešnim leksičkim odabirom kod nekog ključnog koncepta (v. istraživanje koje su proveli Federico i sur. 2014, potpoglavlje 2.3.3.4). Istraživačko pitanje IP4 povezuje metodu analize pogrešaka s metodom subjektivne procjene i osobito je aktualno otkada je u uporabi neuronska paradigma (v. potpoglavlje 2.1.3) jer su izlazni podaci danas sve tečniji i sadržavaju manje morfosintaktičkih pogrešaka, što bi moglo dovesti do boljih ocjena razumljivosti segmenata, čak i ako ne odgovaraju u potpunosti izvornim rečenicama. S druge strane, upravo zato što su leksičke pogreške toliko ključne i u pogledu percepcije prijevoda (v. Simeon 2008, Federico i sur. 2014, Freitag i sur. 2021), koji može biti lažno dobar, kao i u tome što one mogu drastično utjecati na nerazumijevanje, provjerit će se istinitost formiranog istraživačkog pitanja IP5.

Kao što je već navedeno, dosljednost u vrednovanju i visoke razine slaganja vrlo su važne kako bi se rezultati mogli smatrati objektivnima, a promjene u postizanju objektivnih podataka potrebno je pratiti i dijakronijski. Zadnje istraživačko pitanje (IP6) pritom povezuje sve metode odnosno dovodi u suodnos analizu pogrešaka prvo s dosljednošću i slaganjem ispitanika iz metode subjektivne procjene, a potom i s rezultatima metode testa razumijevanja. Time se usmjerava pažnja na pitanje koje trenutačno zabrinjava mnoge istraživače: kako će se metode vrednovanja nositi s povećavanjem općenite kvalitete strojnih prijevoda i hoće li uspjeti biti dovoljno robusne i kako se broj pogrešaka u prijevodima bude smanjivao (v. npr. Freitag i sur. 2021: 1460).

4. Metodologija

Ovo se istraživanje iz područja računalne lingvistike i traduktologije sastoji od četiriju dijelova: dva zadatka u kojima ispitanici određuju kvalitetu zadanog strojno prevedenog materijala (subjektivnom procjenom i testom razumijevanja), anketnog obrasca koji ispituje stavove i prijašnja iskustva ispitanika s alatima za strojno prevođenje te analize pogrešaka koju provodi profesionalni prevoditelj.

4.1 Materijal

Istraživanja strojnoga prevođenja, koja su predstavljena i u poglavlju Teorijski okvir, najčešće se provode s velikim svjetskim jezicima. Pritom svakako prednjači engleski kao *lingua franca* modernog svijeta, a obično ga se dovodi u odnos s nekim drugim većim ili manjim jezikom (v. npr. Federico i sur. 2014 za englesko-kineski, englesko-ruski i englesko-arapski jezični par, ali i npr. Benkova i sur. 2021 za englesko-slovački jezični par ili Bojar 2011 za englesko-češki jezični par). I u Hrvatskoj se u središtu zanimanja najčešće nalazi kombinacija englesko-hrvatski, premda ne isključivo (v. npr. Seljan i sur. 2015 za njemačko-hrvatski i rusko-hrvatski jezični par, uz englesko-hrvatski, ili Ljubas 2018 za švedsko-hrvatski i Ljubas 2020 za njemačko-hrvatski jezični par). Pritom je glavni razlog zašto se krajnji korisnici služe alatima za strojno prevođenje upravo da shvate značenje nekog teksta napisanog na jeziku kojim ne vladaju, što obično uključuje i manje zastupljene jezike, koliko god da su prijevodi s njih uglavnom niže kvalitete (v. Fan i sur. 2021).

U ovom je istraživanju odabran njemačko-hrvatski jezični par. Stoga je za materijal istraživanja izabran tekst koji je izvorno napisan na njemačkom jeziku, a ključan čimbenik pri odabiru teksta bio je da je tekst praćen skupom pripadajućih pitanja prilagođenih provjeri razumijevanja pročitano. Takav je predložak preuzet s mrežnih stranica udruženja njemačkih kulturnih centara Goethe-Institut, koje je certificirano za provedbu standardiziranih ispita znanja njemačkog jezika diljem svijeta. Jezični stručnjaci za provedbu takvih ispita odabiru tekstove i pomno izrađuju setove pitanja kako bi provjerili razinu znanja njemačkog jezika kod kandidata, i to od razine A1 do C2 prema europskom referentnom okviru. Neki od tih profesionalnih materijala za provjeru jezičnih znanja, poput oglednih primjeraka već provedenih ispita, javno su dostupni na mrežnim stranicama Goethe-Instituta kao pomoć

kandidatima u pripremi za pristupanje ispitu. Upravo iz jednog takvog oglednog primjerka ispita, za razinu C2 prema europskom referentnom okviru, preuzet je polazni tekst za ovo istraživanje¹³: kolumna imena „Nicht immer nur klagen“ (*Ne treba se uvijek samo žaliti*, slobodan prijevod) iz prvog dijela ispita koji se odnosi na kompetenciju čitanja s razumijevanjem, kao i pripadajuća pitanja uz tekst. Tekstovi na kojima ispitivači provjeravaju jezično znanje kandidata na ispitima ovakvog tipa moraju ispunjavati nekoliko zahtjeva: važno je da su pisani općim jezikom jer se njima ne ispituju stručna znanja kandidata o nekoj temi, već generalna jezična kompetencija, te je važno da su aktualne tematike ili primjereni za dob kandidata, a u ovom slučaju to su odrasli. Upravo su zato ovakvi korpusi tekstova prikladni i za istraživače koji provjeravaju kvalitetu strojnih prijevoda. Preuzeta pitanja su pitanja višestrukog izbora od kojih je uvijek samo jedan odgovor točan, a izvorno ih je uz tekst bilo priloženo jedanaest.

S obzirom na visoke kognitivne zahtjeve koji se pred ispitanika stavljaju u dužim istraživanjima s nekoliko etapa, tekst je prvo bilo potrebno prilagoditi za ovu svrhu. Skraćen je na dvije kartice odnosno petsto šezdeset i tri riječi, pri čemu se pazilo da se ne naruši koherentnost i unutarnja logika teksta, već da on i dalje funkcionira kao jasna, potpuna cjelina. Valja istaknuti da se tekst nije uređivao s ciljem da bi ga se pripremilo za strojni prijevod, jezik se nije kontrolirao kako bi se olakšalo strojno prevođenje, već se uređivanje isključivo odnosilo na prilagodbu teksta za vrednovanje. Na ovom mjestu prikazan je završni oblik izvornika kakav je postao dijelom istraživanja, podijeljen na trideset i osam rečeničnih segmenata:

1. *Nicht immer nur klagen!*
2. *Ich bin Lehrerin und unterrichte Deutsch und Geschichte.*
3. *Irgendwie ist dieser erste Satz nicht nur die kurze und knappe Information über meinen Beruf, sondern ein Geständnis.*
4. *Der Satz „Ich bin Lehrerin“ kommt mir stets wie ein Bekenntnis vor, wie ein Bekenntnis zu etwas Negativem, ungefähr ähnlich dem Bekenntnis zu einem Laster wie dem Rauchen oder Schlimmerem.*
5. *Wieso eigentlich?*
6. *Ich mag meinen Beruf.*

¹³ Dostupan na: https://bfu.goethe.de/c2_mod/lesen.php (zadnji pristup: 6. lipnja 2021).

7. *Nicht wegen des Beamtenstatus, der mich nicht den Unwägbarkeiten des Arbeitsmarktes aussetzt.*
8. *Der ist zwar angenehm, das gebe ich zu, aber nicht entscheidend.*
9. *Und die viele Freizeit – die ist sowieso nur eine Illusion.*
10. *Nein, als ich damals “auf Lehramt” studierte, steckte ich voller Ideale.*
11. *Die Arbeit mit Kindern und Jugendlichen, die Vermittlung von Wissen und Werten – ich konnte mir keinen besseren Beruf vorstellen.*
12. *Eigentlich müsste ich mich mit meinem Beruf nicht verstecken, könnte stolz auf ihn sein.*
13. *Das bestätigte mir vor Kurzem eine Umfrage, nach der 64 Prozent der Befragten den Lehrern bescheinigten, dass sie gute oder sogar sehr gute Arbeit leisteten.*
14. *Also jubiliere und an die Brust klopfen?*
15. *Da ist aber auch die andere Seite der Medaille.*
16. *Ich glaube, kaum ein Berufsstand wurde in den vergangenen Jahren so stark angegriffen wie der der Lehrer.*
17. *Da sind die diversen Pisa-Studien, die immer wieder belegen, dass unsere Schüler im internationalen Vergleich gerade mal nur Mittelmaß sind.*
18. *Wer trägt dafür die Verantwortung?*
19. *Letztendlich sind es immer die Lehrer, die die Sache nicht in den Griff kriegen, sich zu wenig Mühe geben oder gar für ihren Beruf nicht geeignet sind.*
20. *Was können wir Lehrer tun, um die Dinge zum Besseren zu wenden?*
21. *Natürlich könnten wir immer wieder auf die anderen zeigen, darüber klagen, was uns angetan wird und auf unsere kaum zu bewältigenden Aufgaben verweisen – denn wir müssen nämlich oft auch Erziehungsaufgaben der Eltern übernehmen.*
22. *Wir könnten uns aber auch an die eigene Nase fassen und überlegen, wie wir gegen all das angehen können.*
23. *Es ist schon komisch, dass wir Lehrer, die wir stets und ständig andere und ihre Arbeit bewerten und beurteilen, nicht in der Lage sind, souverän mit Kritik oder Urteilen über uns selber umzugehen.*
24. *Wenn wir Lehrer untereinander ohne Scheu und falsche Scham die Karten auf den Tisch legen, müssen wir alle zugeben, dass unser Alltag von permanenter Angst begleitet wird – der Angst, vor der Klasse zu scheitern.*
25. *Wir müssen zugeben, dass keine Unterrichtsstunde reine Routine ist, sondern das Gegenteil: immer wieder ein Kampf um Anerkennung, ein Ringen um Aufmerksamkeit und Disziplin.*
26. *Und was tun wir dagegen?*

27. *Bereiten Fachlehrer gemeinsam Unterrichtseinheiten vor?*
28. *Probieren wir alternative Methoden aus?*
29. *Entwickeln wir gemeinsam Kommunikationsformen, die uns näher an die Wünsche und Vorstellungen unserer Schüler bringen?*
30. *Leider nicht.*
31. *In der Mehrzahl der Fälle bleibt jeder Kollege der traditionelle Einzelkämpfer und das, was sich hinter der Klassentür abspielt, bleibt geheim.*
32. *Es gibt inzwischen Schulen, an denen eine kontinuierliche Lehrerbewertung durch die Schüler eingeführt wurde.*
33. *Damit sie funktioniert, muss dreierlei passieren: Erstens müssen die bewertenden Schüler anonym bleiben.*
34. *Die Schüler müssen zweitens gelernt haben, wie man Kritik übt.*
35. *Und drittens erfährt allein die betroffene Lehrkraft die Ergebnisse, mit denen sie dann umgehen kann, wie sie will.*
36. *Bisher ist erst ein Viertel der bewerteten Lehrerinnen und Lehrer bereit, mit ihrer jeweiligen Klasse über die geäußerte Kritik zu sprechen.*
37. *Manchen Schülern ist das zu wenig und sie fragen nach dem Sinn der ganzen Frageaktion.*
38. *Ihnen möchte ich antworten: Habt Geduld, immerhin ist ein Anfang gemacht.*

Ovaj je uređeni tekst potom povjeren profesionalnom prevoditelju za njemački jezik, odnosno prvo je konvencionalnim putem preveden na hrvatski jezik, kako bi nastao ljudski prijevod koji će u nastavku istraživanja služiti kao referentni:

1. *Nemojte se stalno žaliti!*
2. *Nastavnica sam i predajem njemački te povijest.*
3. *Nekako mi ta prva rečenica ne zvuči samo kao kratka i jasna informacija o mom zanimanju, već kao priznanje.*
4. *Rečenica „Ja sam nastavnica“ uvijek mi zvuči kao neko priznanje, kao da priznajem nešto negativno, otprilike kao da imam neki porok, recimo da pušim ili što još gore.*
5. *A zašto?*
6. *Volim svoj posao.*
7. *Ne zato što sam zaposlena u državnoj službi, pa zbog toga nisam izložena nepredvidljivosti tržišta rada.*

8. *Priznat ću da je i to ugodno, ali nije najvažnija stvar.*
9. *A sve ono slobodno vrijeme koje se spominje – to je ionako samo iluzija.*
10. *Ne, kada sam upisala studij da postanem nastavnica, bila sam puna ideala.*
11. *Rad s djecom i mladima, prenošenje znanja i vrijednosti – nisam mogla zamisliti bolje zvanje.*
12. *Zapravo uopće ne bih trebala skrivati da sam nastavnica, trebala bih moći biti ponosna na to.*
13. *To mi je nedavno potvrdila i jedna anketa prema kojoj je 64 % ispitanika priznalo da nastavnici dobro ili čak vrlo dobro obavljaju svoj posao.*
14. *Znači, imamo se čemu veseliti?*
15. *Postoji, međutim, i druga strana medalje.*
16. *Mislim da nema te struke koja je proteklih godina bila tako snažno napadnuta kao nastavnička.*
17. *Iskaču neke PISA-ine studije koje neprestano dokazuju da su naši učenici u međunarodnoj usporedbi tek osrednji.*
18. *Tko za to snosi odgovornost?*
19. *Na kraju su uvijek nastavnici ti koji navodno nemaju stvari pod kontrolom, koji se premalo trude ili čak uopće nisu sposobni za obavljanje njihova zanimanja.*
20. *Što mi nastavnici možemo učiniti da okrenemo stvari nabolje?*
21. *Naravno da bismo mogli upirati prstom u druge, žaliti se na sve ono što nam je učinjeno i ukazati na to koliko je teško savladati sve naše zadatke – jer često, naime, moramo preuzeti i odgojnu ulogu roditelja.*
22. *Ali mogli bismo i kritički postaviti sami prema sebi i razmisliti o tome kako da se nosimo sa svime ovime.*
23. *Ipak je smiješno da mi nastavnici, koji neprestano (pr)ocjenjujemo druge i njihov rad, nismo u stanju suvereno se nositi s kritikom ili mišljenjima o nama samima.*
24. *Kada bismo si mi nastavnici bez straha i lažnog srama položili karte na stol, morali bismo priznati da nam je svakodnevnica popraćena neprekidnim strahom – strahom od neuspjeha pred razredom.*
25. *Morali bismo priznati da nijedan nastavni sat nije čista rutina, već upravo suprotno: uvijek je on iznova borba za priznanjem, pažnjom i disciplinom.*
26. *A kako se borimo protiv toga?*
27. *Pripremaju li predmetni nastavnici zajednički lekcije?*
28. *Iskušavamo li alternativne metode?*

29. *Razvijamo li zajednički oblike komunikacije koji bi nas približili željama i predodžbama naših učenika?*
30. *Nažalost, ne.*
31. *U većini slučajeva svaki kolega bori se sam za sebe, a ono što se odvija iza zatvorenih vrata učionice ostaje tajna.*
32. *U međuvremenu postoje škole u kojima je uvedeno kontinuirano vrednovanje nastavnika koje provode učenici.*
33. *Da bi takvo vrednovanje funkcioniralo, moraju se dogoditi tri stvari: Učenici koji vrednuju nastavnike prvenstveno moraju ostati anonimni.*
34. *Učenici zatim moraju naučiti kako izreći kritiku.*
35. *I, treće, samo dotični nastavnik smije saznati rezultate vrednovanja te s njima potom smije raditi što želi.*
36. *Dosad je tek četvrtina vrednovanih nastavnika i nastavnica spremna sa svojim razredom porazgovarati o izrečenoj kritici.*
37. *Neki učenici smatraju da to nije dovoljno i pitaju koji je uopće smisao čitavog upitnika.*
38. *Njima bih htjela odgovoriti: Strpite se, barem se nešto pokrenulo.*

Od jedanaest pitanja koja su bila dostupna za ovaj tekst, kao dio testnog seta u ovom istraživanju odabrano je sedam pitanja. Eliminirana su ona pitanja na koje se odgovor krio u dijelovima teksta koji su bili skraćeni u postupku pripreme. Kako poznavanje njemačkog jezika nije bio preduvjet za sudjelovanje u istraživanju (v. sljedeće potpoglavlje), istom je profesionalnom prevoditelju također bilo povjereno da prevede i set pitanja i odgovora, kako bi se mogli postaviti ispitanicima. Pitanja je naknadno bilo potrebno dodatno urediti kako bi bila što neutralnija: primjerice, u izvornim pitanjima ponavlja se riječ *Autorin* (njem. *autorica*), ali kako je poznato da alati za strojno prevođenje nerijetko griješe kod rodničkih odrednica ili u ovakvim slučajevima u sročnosti imenica i glagola, u prevedenom je pitanju ta riječ zamijenjena neutralnom odrednicom *osoba*, s ciljem da se samim pitanjem ne otkrije informacija koja nužno ne proizlazi iz izlaznih podataka. Konačni oblik pitanja i odgovora testnog seta glase:

- (1) *Kako se osoba osjeća kada govori o svom zanimanju?*
- (A) *Osjeća se vrlo važno.*
 - (B) *Ima pomiješane osjećaje.*
 - (C) *Ljuti se na one koji je to pitaju.*
 - (D) *Ne osjeća ništa.*
- (2) *Zašto se osoba odlučila za nastavničko zanimanje?*
- (A) *Htjela je mlade podučavati stvarima važnima za život.*
 - (B) *Tražila je posao koji joj pruža mnogo slobode.*
 - (C) *Sviđala joj se sigurnost tog radnog mjesta.*
 - (D) *Tražila je odgovoran posao.*
- (3) *Kakav je, prema mišljenju ove osobe, stav javnosti prema nastavničkom zanimanju?*
- (A) *Nastavnike se krivi za sve što je u društvu loše.*
 - (B) *Negativan stav prema nastavnicima više nije toliko izražen kao prije.*
 - (C) *Nastavnici su istovremeno i vrlo priznati i snažno ih se kritizira.*
 - (D) *Nastavnike se uglavnom smatra nesposobnima.*
- (4) *S obzirom na opisanu situaciju, ova se osoba zalaže za to da...*
- (A) *...roditelji preuzmu veću odgovornost za odgoj svoje djece.*
 - (B) *...društvo pruži bolju podršku nastavnicima.*
 - (C) *...se učenicima smanji pritisak i opterećenost uspjesima.*
 - (D) *...nastavnici razmisle o tome kako sami mogu promijeniti neke stvari nabolje.*
- (5) *Osoba smatra da se nastavnici boje...*
- (A) *...da nisu u tijeku s najnovijim znanstvenim i stručnim spoznajama.*
 - (B) *...da ih učenici neće dobro prihvatiti.*
 - (C) *...razgovarati s kolegama o svojim problemima.*
 - (D) *...u nastavi iskušati nešto novo.*
- (6) *Osoba smatra da bi bilo dobro kada bi...*
- (A) *...nastavnici zajedničkim snagama osvježavali didaktičke pristupe u nastavi.*
 - (B) *...nastavnici snažno usmjerili pažnju javnosti na svoje probleme.*
 - (C) *...kolege vrednovali rad svakog nastavnika.*
 - (D) *...se u suradnji s učenicima razvijali novi modeli podučavanja i učenja.*

(7) Prva iskustva s vrednovanjem nastavnika pokazuju da...

- (A) ...su tek rijetki učenici spremni pružiti konstruktivnu kritiku.
- (B) ...učenici brzo izgube volju za sudjelovanjem.
- (C) ...se učenici boje da će se ugroziti njihovo pravo na anonimnost.
- (D) ...tek rijetki nastavnici razgovaraju o kritici koju su dobili od učenika.

Na koncu je još trebalo odabrati alat za strojno prevođenje kojim bi se preveo odabrani tekst. Kako se ovo istraživanje bavi provjerom razumljivosti, što ujedno znači da je orijentirano na krajnje korisnike sustava za strojno prevođenje koji niti moraju nužno biti prevoditelji niti se takvim prijevodima služiti u profesionalne svrhe, odabran je sustav za strojno prevođenje kojim se svakodnevno služi najveći broj korisnika, a koji podržava hrvatski jezik: Google Prevoditelj. Izvornik je njegovom neuronskom inačicom strojno preveden 6. lipnja 2021. godine, čime je dobiven paralelan njemačko-hrvatski tekstni materijal odnosno usporedni segmenti s izvornikom i strojnim prijevodom:

1. *Nemojte se samo žaliti!*
2. *Ja sam učiteljica i predajem njemački jezik i povijest.*
3. *Ova prva rečenica nekako nije samo kratka i jezgrovita informacija o mom poslu, već priznanje.*
4. *Rečenica "Ja sam učitelj" uvijek mi se čini kao priznanje, kao priznanje nečemu negativnom, otprilike slično priznanju poroku poput pušenja ili još gore.*
5. *Zašto?*
6. *Volim svoj posao.*
7. *Ne zbog statusa državnog službenika, koji me ne izlaže neizvjesnostima na tržištu rada.*
8. *Ugodno je, priznat ću, ali nije presudno.*
9. *I svo slobodno vrijeme - to je ionako samo iluzija.*
10. *Ne, kad sam studirao "za učitelja", bio sam pun ideala.*
11. *Rad s djecom i mladima, prenošenje znanja i vrijednosti - nisam mogao smisliti bolji posao.*
12. *Zapravo, ne bih se trebao skrivati sa svojim poslom, mogao bih biti ponosan na njega.*
13. *To je potvrdilo i nedavno istraživanje, prema kojem je 64 posto ispitanih reklo da su učitelji dobro ili čak vrlo dobro radili posao.*
14. *Pa navijati i kucati u prsa?*

15. *Ali postoji i druga strana medalje.*
16. *Mislím da teško da postoji zanimanje koje je posljednjih godina tako ozbiljno napadnuto kao učitelji.*
17. *Postoje razne studije iz Pise, koje više puta dokazuju da su naši studenti u međunarodnoj usporedbi tek osrednji.*
18. *Tko je odgovoran za ovo?*
19. *Na kraju, uvijek su učitelji ti koji ne mogu stvari kontrolirati, koji se ne trude dovoljno ili su čak neprikladni za svoju profesiju.*
20. *Što mi učitelji možemo učiniti da stvari promijenimo na bolje?*
21. *Naravno, mogli bismo uvijek iznova ukazivati na ostale, žaliti se na ono što nam se radi i pozivati se na zadatke kojima teško možemo upravljati - jer često moramo preuzeti i zadatke roditeljstva.*
22. *Ali mogli bismo pogledati i vlastiti nos i razmotriti kako se možemo suprotstaviti gore spomenutoj opreznosti.*
23. *Smiješno je da mi učitelji, koje neprestano i neprestano ocjenjujemo i ocjenjujemo druge i njihov rad, nismo u stanju pouzdano se nositi s kritikama ili prosudbama o sebi.*
24. *Ako mi učitelji stavljamo karte na stol jedni drugima bez sramežljivosti i lažnog srama, svi moramo priznati da je naša svakodnevica popraćena trajnim strahom - strahom od neuspjeha pred razredom.*
25. *Moramo priznati da nijedna lekcija nije čista rutina, već upravo suprotno: uvijek iznova borba za priznanje, borba za pažnju i disciplinu.*
26. *I što mi poduzimamo po tom pitanju?*
27. *Pripremaju li učitelji predmetne nastave zajedno?*
28. *Pokušavamo li alternativne metode?*
29. *Razvijamo li zajednički oblike komunikacije koji nas približavaju željama i idejama naših učenika?*
30. *Nažalost ne.*
31. *U većini slučajeva svaki kolega ostaje tradicionalni usamljeni ratnik, a ono što se događa iza vrata razreda ostaje tajna.*
32. *Sada postoje škole u kojima je uvedeno kontinuirano ocjenjivanje nastavnika od strane učenika.*
33. *Da bi to uspjelo, moraju se dogoditi tri stvari: Prvo, studenti koji ocjenjuju moraju ostati anonimni.*
34. *Drugo, studenti su sigurno naučili kako kritizirati.*

35. *I treće, samo dotični učitelj doživljava rezultate s kojima se onda može nositi kako želi.*
36. *Do sada je samo četvrtina ocijenjenih učitelja spremna razgovarati sa svojim razredom o izrečenim kritikama.*
37. *Nekim studentima to nije dovoljno i pitaju o značenju cijelog postupka ispitivanja.*
38. *Želio bih vam odgovoriti: budite strpljivi, barem je početak započet.*

Ovdje predstavljeni strojni prijevod nije poslije još redigiran, već je u neobrađenom obliku kako je izišao iz alata za strojno prevođenje prepušten ispitanicima na vrednovanje. Za potrebe ranije provedenog predistraživanja strojni je prijevod istog izvornika generiran Google Prevoditeljem i 14. svibnja 2019. godine, a kako je potreban za odgovor na IP6, ovdje se predstavlja u cijelosti:

1. *Nemojte se uvijek žaliti!*
2. *Ja sam učiteljica i podučavam njemački jezik i povijest.*
3. *Nekako, ova prva rečenica nije samo kratka i sažeta informacija o mojoj profesiji, nego i ispovijed.*
4. *"Izraz" Ja sam učitelj "uvijek se čini kao ispovijed, kao obveza prema nečemu negativnom, poput obveze prema poroku pušenje ili još gore.*
5. *Zašto?*
6. *Sviđa mi se moj posao.*
7. *Ne zbog statusa državnih službenika, što me ne izlaže hirovima tržišta rada.*
8. *To je ugodno, priznajem, ali ne i presudno.*
9. *I mnogo slobodnog vremena - to je ionako samo iluzija.*
10. *Ne, kad sam u to vrijeme proučavao "učenje", bio sam pun ideala.*
11. *Rad s djecom i adolescentima, podučavanje znanja i vrijednosti - nisam mogao zamisliti bolji posao.*
12. *Zapravo, ne bih morao skrivati svoj posao, mogao bih biti ponosan na njega.*
13. *To je nedavno potvrdila i anketa u kojoj je 64 posto ispitanika izjavilo da je radilo dobro ili čak vrlo dobro.*
14. *Tako slaviti i kucati na grudi?*
15. *Ali postoji i druga strana medalje.*
16. *Vjerujem da je u posljednjih nekoliko godina kao učiteljica gotovo da nije bilo tako strogo napadnuto zanimanje.*

17. *Postoje razne studije u Pisi koje dokazuju da su njemački studenti samo prosječni u međunarodnoj usporedbi.*
18. *Tko snosi odgovornost za to?*
19. *Uostalom, uvijek su nastavnici ti koji ne mogu kontrolirati stvari, koji ne ulažu dovoljno truda u svoj posao, ili koji nisu sposobni za svoj posao.*
20. *Što mi nastavnici možemo učiniti da promijenimo stvari na bolje?*
21. *Naravno, uvijek bismo mogli ukazati na druge, žaliti se na ono što nam je učinjeno i odnositi se na teško upravljive zadatke - jer često moramo preuzeti poslove roditeljstva.*
22. *Ali također možemo uzeti svoj nos i razmisliti o tome kako se možemo pozabaviti spomenutim Duckmäusertumom.*
23. *Smiješno je da mi učitelji koji neprestano i stalno procjenjujemo i vrednujemo druge i njihov rad nisu u poziciji da se pouzdano pozabave kritikom ili prosudbama o nama samima.*
24. *Ako mi učitelji položimo jedni druge na stol bez sramežljivosti i lažne sramote, svi moramo priznati da je naš svakodnevni život popraćen stalnim strahom - strahom od neuspjeha pred razredom.*
25. *Moramo priznati da nijedna lekcija nije čista rutina, već suprotno: uvijek borba za priznanje, borba za pažnju i disciplinu.*
26. *I što ćemo učiniti?*
27. *Jesu li predmetni nastavnici zajedno pripremili nastavne jedinice?*
28. *Pokušavamo li alternativne metode?*
29. *Da li zajedno razvijamo oblike komunikacije koji nas približavaju željama i idejama naših učenika?*
30. *Nažalost ne.*
31. *U većini slučajeva, svaki kolega ostaje tradicionalni usamljeni borac, a ono što se događa iza učionice ostaje tajna.*
32. *Sada postoje škole u kojima su studenti provodili kontinuirano ocjenjivanje nastavnika.*
33. *Postoje tri stvari koje ga čine uspješnim: Prvo, učenici koji ocjenjuju moraju ostati anonimni.*
34. *Drugo, učenici su morali naučiti kritizirati.*
35. *I treće, samo učitelj o kojem se radi uči rezultate, koje oni mogu podnijeti kako žele.*
36. *Do sada je samo četvrtina ispitanih nastavnika spremna razgovarati sa svojim razredom o kritikama.*
37. *Ovo nije dovoljno za neke učenike i oni pitaju za značenje cijele reakcije pitanja.*
38. *Želim odgovoriti: Imajte strpljenja, nakon svega, napravljen je početak.*

U Prilogu A usporedno se mogu pratiti izvornik, izlazni podaci iz 2019. godine, izlazni podaci iz 2021. godine te referentni prijevod.

4.2 Ispitanici

Pregledom relevantne literature utvrđeno je da na rezultate istraživanja uvelike utječe odabir skupine ispitanika – ispitanici mogu biti laici ili profesionalni prevoditelji, a također treba donijeti izbor između dvojezičnih ispitanika ili onih koji ne govore izvorni jezik testnog materijala, kao i odlučiti koliko će homogena biti skupina ispitanika. U preliminarnom istraživanju odabrana je relativno heterogena skupina ispitanika: strojni prijevod istog izvornog teksta, generiran Google Prevoditeljem 14. svibnja 2019. godine, vrednovalo je četrdeset i petero ispitanika među kojima je bilo i profesionalnih redaktora strojnih prijevoda te prevoditelja koji se strojnim prevođenjem služe u svakodnevnom radu, ali i ispitanika koji se sa strojnim prevođenjem prije nisu susreli, ili pak tek povremeno u privatnom životu. Također su neki ispitanici bili dvojezični, a drugi ne. Rezultati tog predistraživanja (Ljubas 2019) ukazali su na važnost što homogenije skupine ispitanika kad su u pitanju iskustava sa strojnim prevođenjem jer je u razgovoru s ispitanicima uočeno da su vrednovanju pristupili imajući na umu posve različite uporabe strojnih prijevoda. Metodološki gledano, uočeno je i da je važno da nitko od ispitanika ne govori jezikom izvornika – u predistraživanju su ispitanici sa znanjem njemačkog bili u prednosti nad onim ispitanicima koji ga ne govore jer su sa svojim jezičnim znanjima lakše mogli proniknuti značenje u dijelovima teksta koji su, primjerice, doslovno prevedeni.

S obzirom na to da se u predmetnom istraživanju vrednuje kriterij razumljivosti izlaznih podataka, odabrani su ispitanici koji se strojnim prijevodima ne služe u prevoditeljske svrhe i ispitanici koji kao prevoditeljski laici nemaju iskustva s redigiranjem strojnih prijevoda. Time je osigurano da ispitanici istraživanju pristupaju s ciljem asimilacije podataka na umu, dakle sa svrhom izvlačenja informacija iz tekstova napisanih na jezicima kojima inače ne vladaju. Ispitanici također nisu prošli nikakvu prethodnu obuku o vrednovanju strojnog prevođenja jer se željelo ispitati njihovo laičko mišljenje.

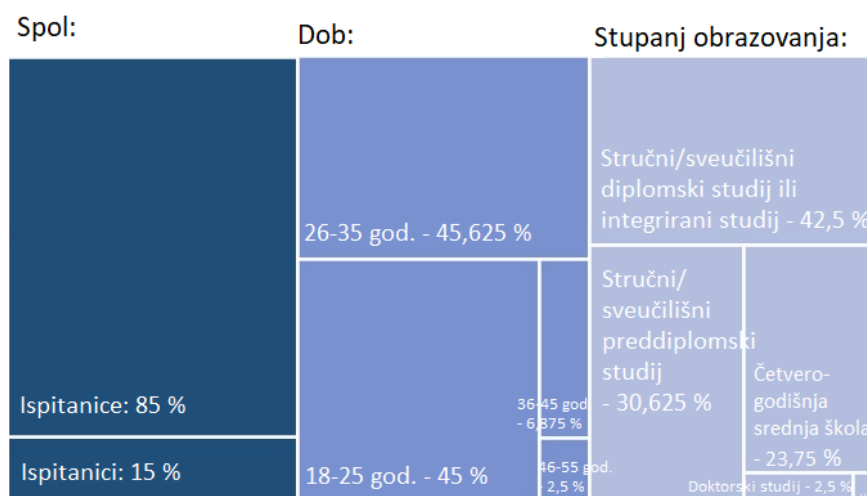
Također su odabrani jednojezični ispitanici, što se u ovom kontekstu odnosi ili na ispitanike koji uopće ne vladaju njemačkim jezikom ili na ispitanike koji eventualno poznaju osnove njemačkog jezika, ali se njime niti služe aktivno niti imaju dovoljna pasivna znanja da bi bili u

stanju razumijevati čitave tekstove. Riječ je o ispitanicima koji bi se hipotetski i neovisno o istraživanju mogli poslužiti strojnim prevođenjem da shvate smisao neke njemačke kolumne poput ove. Neka su istraživanja pokazala da se od skupine jednojezičnih ispitanika može očekivati veća razina slaganja nego od skupine dvojezičnih ispitanika (v. Guzmán i sur. 2015), a velika je prednost rada s njima i njihova rasprostranjenost i dostupnost.

Primarni je nedostatak rada s jednojezičnim ispitanicima najčešće taj što je u istraživanje potrebno uključiti referentni prijevod, kojemu ispitanici pristupaju kao savršenom prijevodu unatoč pogreškama koje se i u njemu mogu potkrasti. Međutim, u ovom je istraživanju taj nedostatak eliminiran jer je koncipirano na način da ispitanicima uopće nije potreban referentni prijevod – oni ne vrednuju točnost, i stoga ne moraju uspoređivati prijevod s izvornikom ili referentnim prijevodom, pretražujući jesu li doista baš sve informacije vjerno prenesene.

U istraživanju je sudjelovalo ukupno sto šezdeset ispitanika, a na Slici 3 slikovito je prikazan pregled ispitanika grupiranih po prikupljenim sociodemografskim podacima. Od sto šezdeset ispitanika 15 % činili su muškarci, a preostalih 85 % žene. Ne smatra se da je ova nejednaka distribucija ispitanika mogla utjecati na rezultate istraživanja jer tema disertacije nije specifično vezana za određeni spol niti rod, te ne postoje inherentne razlike između ispitanika i ispitanica u pogledu sposobnosti vrednovanja strojnih prijevoda. Upravo stoga što se kategorija spola u ovim istraživanjima smatra neutjecajnom, pitanje koje se odnosilo na spol ispitanika nije bilo obavezno ispuniti, odnosno svi su ispitanici mogli slobodno nastaviti s istraživanjem i ako se o spolu uopće ne izjasne. Ipak, svi su ispitanici koji su pristupili istraživanju svojevrijedno odgovorili na pitanje.

Sociodemografska obilježja ispitanika



Slika 3. Pregled sociodemografskih obilježja ispitanika.

Ispitanici su morali imati najmanje osamnaest godina kako bi mogli pristupiti istraživanju: sedamdeset i dvoje ispitanika imali su do dvadeset i pet godina, a sedamdeset i troje do trideset i pet godina. Jedanaestero ispitanika bilo je u dobi od trideset i pet do četrdeset i pet godina, a preostalih četvero imali su do pedeset i pet godina. U istraživanju nisu sudjelovali ispitanici stariji od pedeset i pet godina. Kao ni po pitanju spola, to što su ispitanici bili mahom nešto mlađe dobi ne smatra se utjecajnim za njihovu sposobnost vrednovanja, a može se argumentirati i da distribucija ispitanika u nekoj mjeri zrcali stvarnu distribuciju korisnika sustava za strojno prevođenje odnosno može se pretpostaviti da češći korisnici jesu nešto mlađe dobi.

Istraživanju također nije pristupio nijedan ispitanik s osnovnom školom kao zadnjim završenim stupnjem obrazovanja, ali mu je pristupio jedan ispitanik sa završenom trogodišnjom srednjom školom te trideset i osmero ispitanika sa završenom četverogodišnjom srednjom školom. Četrdeset i devetero ispitanika završilo je stručni ili sveučilišni preddiplomski studij, a šezdeset i osmero ispitanika stručni ili sveučilišni diplomski studij ili integrirani studij. Preostalih četvero ispitanika ima završen doktorski studij. Ispitanici su prihvatljivo i ravnomjerno raslojeni po stupnjevima obrazovanja jer se u ispitivanju željelo uključiti realne i potencijalne korisnike strojnih prijevoda u svrhu asimilacije podataka.

4.3 Metoda

Kao što je već navedeno, ovo se istraživanje sastoji od tri svojevrsna zadatka, koji su opisani u nastavku u ovog poglavlja. U dva prva zadatka sudjeluju navedeni ispitanici, a u trećem zadatku odabran je jedan stručni anotator. Podaci su prikupljeni obrascem¹⁴ podijeljenim na dva zadatka, upitnikom o stavovima i dijelom u kojemu se ispituju demografski podaci ispitanika, a u cijelosti je dostupan u Prilogu B. Dobiveni su podaci prvenstveno kvalitativno analizirani.

Prvi se zadatak odnosi na metodu subjektivne procjene kvalitete izlaznih podataka i ispitanici se u njemu dijele u dvije skupine. Od ukupno sto šezdeset ispitanika osamdeset je nasumično odabranih ispitanika činilo skupinu koja je vrednovala raščlanjene segmente strojnoga prijevoda. Svaki segment činila je jedna rečenica prijevoda, što je ukupno tvorilo trideset i osam

¹⁴ Kako je istraživanje provedeno u vrijeme globalne pandemije bolesti COVID-19, podaci nisu prikupljeni uživo sa svih sto šezdeset ispitanika, nego se postupak pojednostavio tako što su zadaci za ispitanike postavljeni unutar mrežnog obrasca i mrežnim su putem distribuirani. Doktorandica je bila na raspolaganju ispitanicima, a imali su i detaljne upute za rješavanje zadataka. To je omogućilo dovoljno brzo, efikasno i, što je najvažnije, za sve sudionike sigurno provođenje istraživanja.

segmenata. U tekstnom materijalu gotovo su jednako zastupljene kratke rečenice (od najmanje dviju do devet riječi), srednje duge rečenice (od deset do osamnaest riječi) i nešto dulje rečenice (od devetnaest do najviše trideset i pet riječi), što je važan aspekt u vrednovanju jer se smatra da rečenična duljina utječe na uspješnost prijevoda, osobito kod neuronskih sustava za strojno prevođenje. Segmenti su ispitanicima predstavljeni nasumičnim redoslijedom, tako da se oni nisu čitali kao koherentni tekst. To je uobičajeni postupak u vrednovanju ovom metodom kojom se postiže da se ispitanici usredotoče na svaki pojedini segment, umjesto da značenja pokušavaju iščitavati iz cjeline. Uputa na početku zadatka glasila je da pročitaju svaki pojedini segment i procijene koliko im je on razumljiv.

Svoje mišljenje ispitanici iskazuju pridružujući svakom segmentu po jednu ocjenu na ljestvici s pet bodova od 1 do 5, gdje 1 označava da je prijevod „nerazumljiv“, 2 da je prijevod „slabo razumljiv“, 3 da je „prilično razumljiv“, 4 da je „dobro razumljiv“ i 5 da je „potpuno razumljiv“. Podaci su analizirani pojedinačno za svaki segment, ali izračunom aritmetičke sredine svih ocjena pridruženih segmentima dobiva se i ocjena razumljivosti svakog ispitanika za cjelokupni tekst, što je podatak koji se može usporediti s rezultatima druge skupine, koja se također sastojala od osamdeset preostalih, nasumično raspoređenih ispitanika. Njima je u prvom zadatku prikazan cijeli strojni prijevod odjednom. Njihov zadatak glasilo je da pročitaju čitav tekst i razmisle o tome koliko im je razumljiv, te da ga potom ocijene na istoj ljestvici kao i prethodna skupina ispitanika. Pri analizi podataka prvenstveno je bilo važno uočiti je li kod jedne skupine postignuta veća razina slaganja i dosljednosti u odnosu na drugu.

Po rješavanju prvog zadatka obje se skupine ispitanika preusmjeravaju na isti drugi zadatak koji se sastoji od testa razumijevanja. Ispitanici moraju pročitati strojni prijevod te potom odgovoriti na sedam pitanja s ponuđenim četiri odgovora od kojih je samo jedan točan. Čitavo to vrijeme na raspolaganju im je cjelokupni strojni prijevod, a isto tako ga imaju vremena pročitati koliko god puta žele i mogu se vraćati na njega dok odgovaraju na pitanja. Ponuđeni odgovori na pitanja napisani su tako da su, logički gledano, jednako vjerojatni, što znači da se u odgovaranju nije moguće oslanjati na opća znanja o svijetu, već da je rješenja potrebno prepoznati u tekstu. Najčešće je za to potrebno povezati nekoliko informacija s različitim mjestima u tekstu, čime se želi ispitati dubinsko razumijevanje prijevoda.

Budući da uvijek preostaje mogućnost da će ispitanici jednostavno pogoditi točan odgovor ili da će neki ispitanici uložiti više truda i kognitivnog napora za rješavanje testa razumijevanja,

uvedena su dva kontrolna elementa. Prvo su na kraju seta pitanja ispitanici zamoljeni da procijene koliko su sigurni u odabrane odgovore, s ponuđenim opcijama:

- a) *Nimalo nisam siguran/na u odabir odgovora.*
- b) *Pretpostavljam da sam odabrao/la većinu netočnih odgovora.*
- c) *Niti sam siguran/na niti nisam.*
- d) *Pretpostavljam da sam odabrao/la većinu točnih odgovora.*
- e) *Sasvim sam siguran/na u odabir odgovora.*

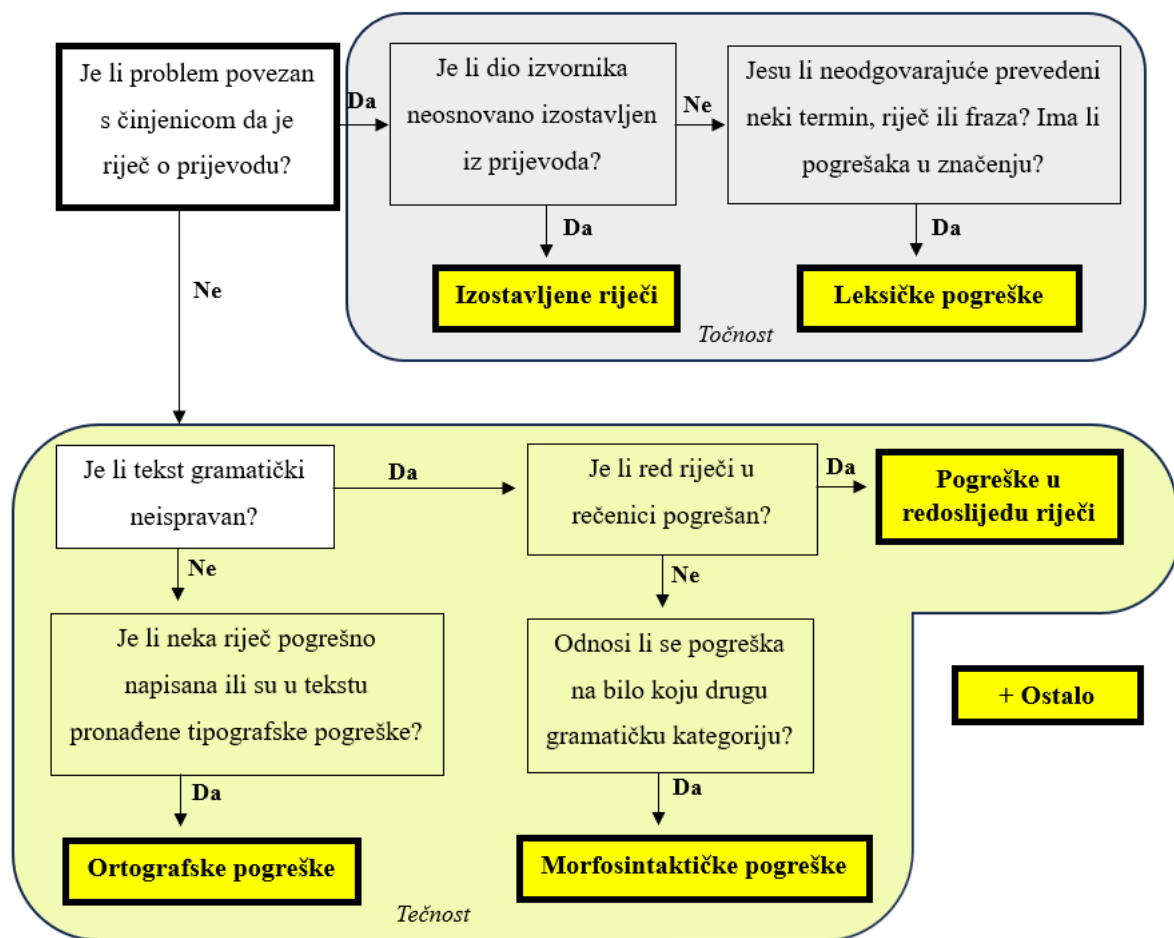
Drugi je kontrolni element vrijeme potrebno za rješavanje testa razumijevanja. Ispitanici su na početku zadatka dobili uputu da mjere vrijeme i na kraju ga upišu u za to predviđenu kućicu. Ti rezultati nadopunjuju kvalitativno tumačenje podataka o uspješnosti na testu. Ovom se metodom iz broja točnih odgovora na testu procjenjuje koliko je neki tekst razumljivo preveden, a može se procijeniti i jesu li neki dijelovi teksta jasnije prevedeni ili gdje je stroj imao više poteškoća.

Nakon testa razumijevanja obje skupine ispitanika pristupaju anketnom obrascu o njihovim stavovima o strojnom prevođenju. Predstavljene su im izjave koje izražavaju pozitivne ili negativne stavove o strojnom prevođenju ili se odnose na to koliko se u životu služe strojnim prevođenjem. Ispitanici stupanj slaganja s predstavljenim izjavama trebaju iskazati na Likertovoj petobodnoj ljestvici, nakon čega su njihovi odgovori kodirani kao pozitivni, neutralni ili negativni. Ti se podaci u analizi uspoređuju sa subjektivnim ocjenama razumljivosti iz prvog zadatka i brojem točnih odgovora na testu razumijevanja, kako bi se provjerila objektivnost rezultata tih dviju metoda, odnosno stupanj ovisnosti rezultata o unaprijed određenim stavovima. Zaključci o stavovima ispitanika izvođeni su isključivo prema kodiranim odgovorima ispitanika, ali diskusija je mjestimično proširena i nekim izdvojenim komentarima ispitanika. Ispitanicima je, naime, pružena mogućnost da pred kraj svog sudjelovanja u istraživanju u otvorenom polju slobodno izraze svoje misli i stavove o strojnom prevođenju. Kako su se „razgovori“ s ispitanicima zbog globalne pandemije provodili putem mrežnog obrasca, to je polje zamijenilo neobavezni razgovor ispitanika s istraživačem. Trideset i četvero je ispitanika iskoristilo mogućnost da dodatno pojašni svoje stavove, pa je dio odgovora predstavljen u radu i služi kao dodatni komentar i mogućnost da se upotpune zaključci dobiveni

osnovnom analizom stavova ispitanika. Na samom kraju istraživanja ispitanici još pristupaju upitniku o demografskim podacima.

Zadnji dio istraživanja čini analiza pogrešaka u izlaznim podacima. Analiza pogrešaka kognitivno je vrlo zahtjevna i zamorna metoda za koju je potrebno imati čitav niz lingvističkih i prevoditeljskih predznanja. Ona se stoga načelno ne prepušta laicima – u ovom bi slučaju bilo prezamorno podučiti svakog pojedinog ispitanika pravilima analize pogrešaka, već je za analizu pogrešaka i u predistraživanju i u glavnom istraživanju (u razmaku od dvije godine) angažiran isti dvojezični ispitanik, po struci diplomirani germanist i lingvist, profesionalni prevoditelj s njemačkog jezika, koji već ima potrebne teorijske kvalifikacije te prethodno iskustvo rada sa strojnim prijevodima. Vrlo je važno bilo angažirati vanjskog anotatora, premda sam ja kao glavna istraživačica s potrebnim kompetencijama u načelu i sama mogla provesti metodu analize pogrešaka. No, to bi značilo da su podaci o metodi vrednovanja analize pogrešaka prikupljeni introspektivno; pritom se i za treći zadatak htio zadržati istovrijedan način analize rezultata kao i u prva dva zadatka kada su odabrani vanjski ispitanici, i gdje ja kao provoditeljica istraživanja zadržavam neutralan položaj i ekstrospektivno pristupam tumačenju rezultata istraživanja.

Uz anotatora za analizu pogrešaka vrlo je važno odabrati pravilnu klasifikaciju pogrešaka. Metrika *Multidimensional Quality Metrics* (MQM) predlaže niz mogućih kategorija pogrešaka, no važno je pripaziti da se ne odabere previše nepotrebnih ili suviše detaljnih kategorija (da proces ne bi postao pretjerano zamoran i rezultati, kao posljedica toga, nedosljedni) ili premalo kategorija (jer podaci na kraju ne bi pružali dovoljno jasne informacije o jezičnoj kvaliteti teksta). U predistraživanju provedena je i analiza pogrešaka, i na materijalu iz 2019. godine utvrđeno je da se najveći broj pogrešaka može klasificirati u šest kategorija: izostavljene riječi, leksičke, morfosintaktičke i ortografske pogreške, pogreške u redosljedu riječi te u kategoriju „ostalo“ koja bi obuhvaćala sve druge uočene pogreške (Ljubas 2019). S odabranim je anotatorom pogrešaka dogovoreno da će se te pogreške zadržati i u glavnom dijelu istraživanja, a pružanje okvira unutar kojega bi razvrstao uočene pogreške trebalo je olakšati anotiranje, tj. umanjiti napore koje bi morao dodatno uložiti u ponovno samostalno osmišljavanje primjrene klasifikacije. Slijedeći primjer mentalnih mapa predloženih u „Praktičnim smjernicama za uporabu MQM-a u znanstvenim istraživanjima o kvaliteti prijevoda“ (v. str. 49) i za ovo je istraživanje složena, i na Slici 4 prikazana, mentalna mapa odabranih kategorija s pripadajućim pitanjima koja pomažu anotatoru u određivanju kategorija:



Slika 4. Prilagođena mentalna mapa za određivanje kategorije pogreške.

Anotator, dakle, polazi od pitanja „Je li problem povezan s činjenicom da je riječ o prijevodu?“ te potom prelazi na kategorije pogrešaka koje, ako je doista riječ o prijevodnom problemu, pripadaju domeni točnosti, ili na kategorije pogrešaka koje se odnose samo na tečnost ciljnog teksta. U domeni točnosti zadržane su dvije kategorije: „izostavljene riječi“, ako nešto u ciljnom tekstu nedostaje, te kategorija nazvana „leksičke pogreške“. Ta kategorija uključuje pogrešno preneseno značenje odnosno neodgovarajuće prevedene termine, riječi i fraze, te objedinjuje kategorije *Terminology*, *Mistranslation* i *Accuracy (general)* izvorne MQM-ove klasifikacije pogrešaka. Kategorija se ovdje naziva „leksičkom“ jer bi semantičke ili značenjske pogreške predstavljale ponešto preširok koncept – često se argumentira da se većina pogrešaka može objasniti semantikom, primjerice: pogrešno glagolsko vrijeme, koje bi se u ovoj klasifikaciji podvelo pod kategoriju „morfosintaktičkih pogrešaka“ ujedno je i pogreška semantičke prirode jer ne odražava točno značenje izvornika. Međutim, kako za gramatičke pogreške postoji

posebna izdvojena kategorija, njih ipak treba svrstati u nju, a kategorija „leksičkih pogrešaka“ smatra se užom kategorijom točnosti.

Druga grana pogrešaka odnosi se na tečnost prijevoda i počinje s razlučivanjem je li tekst gramatički neispravan. Ako se pogreška ne manifestira u gramatici, onda sljedeće pitanje anotatora upućuje na kategoriju „ortografskih pogrešaka“ koja uključuje i pogreške u pravopisu i pogreške u uporabi interpunkcijskih znakova i druge tipografske pogreške. Što se gramatičkih pogrešaka tiče, razlikuju se „pogreške u redosljedu riječi“, koje su u strojnim prijevodima vrlo učestalo zabilježene, te „morfosintaktičke pogreške“ u koje se ubrajaju sve preostale gramatičke pogreške.

Osim navedenih kategorija, izdvojena je i šesta kategorija „ostalo“. Neka pitanja u mentalnoj mapi nemaju predviđene i potvrdne i negativne odgovore. Kada anotator uoči pogrešku koja ne pripada nijednoj od izdvojenih kategorija, onda je može smjestiti u tu slobodnu kategoriju, uz kratki opis i objašnjenje svog tumačenja pogreške.

Radi još lakšeg razumijevanja klasifikacije u nastavku slijede primjeri segmenata u kojima su anotirane pogreške klasificirane u svaku od ovih šest kategorija. Primjeri su preuzeti iz predistraživanja, a uza svaki segment navedeni su i izvornik i referentni prijevod (a tako će primjeri biti prikazani i u nastavku rada, u prikazu rezultata i diskusiji). Iako oni služe kao pomoć u čitanju i razumijevanju primjera u radu, valja podsjetiti da ni izvornik ni referentni prijevod nisu bili na raspolaganju ispitanicima dok su sudjelovali u istraživanju. Prvu kategoriju pogrešaka čine izostavljene riječi:

A. Kategorija pogrešaka – izostavljene riječi:

Izlazni podaci: *To ø je nedavno potvrdila i ø anketa u kojoj je 64 posto ispitanika izjavilo da je ø radilo dobro ili čak vrlo dobro.*

Referentni prijevod: *To mi je nedavno potvrdila i jedna anketa prema kojoj je 64 % ispitanika priznalo da nastavnici dobro ili čak vrlo dobro obavljaju svoj posao.*

Izvornik: *Das bestätigte mir vor Kurzem eine Umfrage, nach der 64 Prozent der Befragten den Lehrern bescheinigten, dass sie gute oder sogar sehr gute Arbeit leisteten.*

Iz ovog je segmenta izostavljeno nekoliko riječi (označeno znakom ø, na ovom i drugim mjestima), ali nisu sve ubrojene među izostavljene pogreške. U izlaznim podacima nedostaje ekvivalent za njemačku riječ *mir*: iz referentnog je prijevoda jasno da autorica teksta piše da je anketa *njoj* nešto potvrdila („To **mi** je nedavno potvrdila...“), umjesto da se to odnosi na neku općenitu potvrdu. Anotator je to ubrojio kao pogrešku, premda je ovdje riječ o lakšem propustu.

Druga riječ koja nedostaje u izlaznim podacima ekvivalent je za riječ *eine* u izrazu *eine Umfrage*. Anotator to ne smatra pogreškom, premda se u referentnom ljudskom prijevodu pojavio puni izraz *jedna anketa*. Moguće je da bi koji drugi anotator i to smatrao pogreškom, ali anotiranje je pogrešaka posao prepušten osjećaju odabranog anotatora. U ovom primjeru odabrani anotator smatra da je izostavljanje riječi *eine* legitimna varijacija u prijevodu te da je značenje i na ovaj način vjerno preneseno. Iz ovoga je primjera odmah vidljivo i da je referentni prijevod tek jedan u nizu mogućih i legitimnih prijevoda izvornog teksta.

U segmentu su izostavljene i riječi koje čine grublje materijalne pogreške zbog kojih je došlo do promjene u značenju čitavog segmenta. Izvorna njemačka rečenica prenosi značenje da su se ispitanici predmetne ankete izjasnili o tome da upravo nastavnici dobro obavljaju svoj posao, no kako je ta riječ izostavljena u izlaznim podacima, rečenica se čita kao da ispitanici sami za sebe izjavljuju da oni dobro obavljaju svoje poslove. Pogreške koje čitatelja navode na pogrešno shvaćanje značenja svakako je važno svrstati u prikladnu kategoriju, a u ovom je slučaju anotator pogrešku pridružio kategoriji izostavljenih riječi.

B. Kategorija pogrešaka – leksičke pogreške:

Izlazni podaci: *Ovo nije dovoljno za neke učenike i oni pitaju za značenje cijele reakcije pitanja.*

Referentni prijevod: *Neki učenici smatraju da to nije dovoljno i pitaju koji je uopće smisao čitavog upitnika.*

Izvornik: *Manchen Schülern ist das zu wenig und sie fragen nach dem Sinn der ganzen Frageaktion.*

Leksičke pogreške odnose se najčešće na pogrešan prijevod punoznačnica, zbog čega te pogreške mogu snažno utjecati na nerazumijevanje čitavog segmenta ili teksta. U ovom primjeru pogrešno je prevedena njemačka složenica *Frageaktion*: alat za strojno prevođenje

izbacio je dvije imenice, *reakcija pitanja*, čije značenje nije jasno. U referentnom prijevodu odabran je ekvivalent *upitnik* koji svakako bolje odgovara značenju njemačke riječi i prikladan je kontekstu. Anotator je ovo računao kao jednu leksičku pogrešku, vodeći se time da se radi o jednom značenjskom konceptu i jednoj riječi u izvorniku, unatoč tome što je pogrešno prevedena dvjema riječima. Pogreške se ne moraju uvijek realizirati na razini riječi, kao što je vidljivo i u sljedećim kategorijama.

C. Kategorija pogrešaka – **morfosintaktičke pogreške:**

Izlazni podaci: *I što ćemo učiniti?*

Referentni prijevod: *A kako se borimo protiv toga?*

Izvornik: *Und was **tun** wir dagegen?*

U morfosintaktičke se pogreške ubrajaju pogrešni glagolski oblici (vrijeme, način, stanje), pogrešna uporaba oblika i vrsta riječi, pogreške u sročnosti ili uporabi prijedloga itd. U navedenom primjeru njemački je prezent u hrvatskom preveden futurom prvim, što u određenim situacijama može biti pravilan odabir, ali u ovom ga kontekstu anotator nije smatrao prikladnim rješenjem.

D. Kategorija pogrešaka – **ortografske pogreške:**

Izlazni podaci: *”Izraz“ Ja sam učitelj ”uvijek se čini kao ispovijed, kao obveza prema nečemu negativnom, poput obveze prema poroku pušenje ili još gore.*

Referentni prijevod: *Rečenica „Ja sam nastavnica“ uvijek mi zvuči kao neko priznanje, kao da priznajem nešto negativno, otprilike kao da imam neki porok, recimo da pušim ili što još gore.*

Izvornik: *Der Satz „Ich bin Lehrerin“ kommt mir stets wie ein Bekenntnis vor, wie ein Bekenntnis zu etwas Negativem, ungefähr ähnlich dem Bekenntnis zu einem Laster wie dem Rauchen oder Schlimmerem.*

Iako je u ovom segmentu zabilježen veći broj različitih pogrešaka, već se na prvi pogled zamjećuje da je došlo do ortografske pogreške povezane s uporabom navodnika. Osim što ne

prate pravila hrvatskog jezika, gdje se znak koji otvara navod postavlja dolje, u izlaznim se podacima nasumično pojavio i treći znak navoda koji ništa ne zatvara. Ovakve pogreške nisu neuobičajene u izlaznim podacima sustava za strojno prevođenje, a osim njih u ovu se kategoriju ubrajaju i pogrešno postavljeni zarezi ili drugi interpunkcijski znakovi.

E. Kategorija pogrešaka – **pogrešan redoslijed riječi:**

Izlazni podaci: *Vjerujem da je u posljednjih nekoliko godina **kao učiteljica** gotovo da nije bilo tako strogo napadnuto zanimanje.*

Referentni prijevod: *Mislim da nema te struke koja je proteklih godina bila tako snažno napadnuta **kao nastavnička.***

Izvornik: *Ich glaube, kaum ein Berufsstand wurde in den vergangenen Jahren so stark angegriffen **wie der der Lehrer.***

Redoslijed riječi izdvojen je kao zasebna kategorija jer se smatra da je to slaba točka sustava za strojno prevođenje. U ovom segmentu zabilježeno je više različitih vrsta pogrešaka, osobito u dvojbenom prijevodnom odabiru „kao učiteljica“, ali upravo se kod tog elementa pojavio i manjkav redoslijed riječi. Taj je dio rečenice sintaktički povezan s „napadnutim zanimanjem“ i ne može se nalaziti odvojeno od njega. Iz ovog je primjera također razvidno da se jedan element zapravo može svrstati u nekoliko kategorija pogrešaka odnosno da riječ nije isključiva jedinica pogreške – u teoriji može biti uočeno više pogrešaka nego što u tekstu ima riječi, a katkad se pogreška može protegnuti i na cjelokupni ton ili stil teksta, manifestira se u dosljednosti prevođenja termina tijekom čitavog teksta, ili bilo koji drugi element koji nadilazi razinu riječi ili rečenice.

F. Kategorija pogrešaka – **ostalo:**

Izlazni podaci: *Tako slaviti i **kucati na grudi?***

Referentni prijevod: *Znači, **imamo se čemu veseliti?***

Izvornik: *Also jublieren und **an die Brust klopfen?***

Zadnja kategorija pogrešaka neophodna je kako bi se njome mogle obuhvatiti sve one ostale, pojedinačne pogreške koje se ne uklapaju ni u jednu drugu kategoriju. Katkad se narav takvih pogrešaka djelomično preklapa s nekom drugom kategorijom, ali ih anotatori zbog njihovih specifičnosti mogu izdvojiti i u zasebnu. Ovdje navedeni primjer anotator nije uključio među leksičke pogreške, uz objašnjenje da su same sastavnice njemačke fraze *an die Brust klopfen* dobro prevedene, ali da je stroj trebao prepoznati da je riječ o frazemu – u prijevodu se zapravo bilo potrebno poslužiti drugim leksičkim materijalom da se prenese smisao rečenice, kao što je to učinjeno u referentnom prijevodu. Usto se u ovu kategoriju vrlo često znaju svrstati cijeli segmenti kada su toliko puni pogrešaka da ih je gotovo nemoguće razdvojiti i razlikovati, odnosno kada nije moguće odrediti gdje koja pogreška počinje, a gdje koja završava. To je bio osobito čest slučaj u prijašnjim fazama vrednovanja sustava za strojno prevođenje.

Gledajući iz lingvističke perspektive, moglo bi se pomisliti da bi analiza pogrešaka zahtijevala veću distinkciju odnosno da su navedene kategorije pogrešaka preširoke. Leksičke, morfosintaktičke i ortografske pogreške zacijelo bi se mogle podijeliti na manje potkategorije, a kategorija „ostalo“ na prvi se pogled može činiti nepreciznom. Tu se svakako valja prisjetiti Burchardtove i Lommelove mentalne mape (Slika 2) i „Praktičnih smjernica za uporabu MQM-a u znanstvenim istraživanjima o kvaliteti prijevoda“ (v. poglavlje 2.3.3.3) koje govore kako nije potrebno, dapače, nije poželjno da se metrike razgranaju na što veći broj kategorija pogrešaka, već da je najbolje da se istraživanje svede na šest-sedam kategorija (Burchardt i Lommel 2014: 5). Osim što nisu sve kategorije uvijek važne za sva istraživačka pitanja, analize pogrešaka uvijek su vrlo zamorni procesi koji izmaraju anotatore i većim se brojem kategorija narušava mogućnost da anotatori pri analizi ostanu dosljedni. Njima se treba pokušati olakšati postupak analize na temelju podataka iz prijašnjih istraživanja – u ovom slučaju znamo da je „pogrešan redoslijed riječi“ problematična kategorija za sustave za strojno prevođenje, napose kada jezični par uključuje slavenske jezike, i nije potrebno da to bude potkategorija drugih sintaktičkih pogrešaka, nego je važno da se ispitanik koncentrira na tu određenu pogrešku. Kategorija „ostalo“ pomaže anotatoru uštedjeti vrijeme kako ne bi mogao baratati čitavim nizom specifičnih kategorija u koje na koncu ne bi ubrojio nijednu pogrešku ili možda tek jednu – a pribrajajući ih u kategoriju „ostalo“ uvijek može ostaviti dodatnu natuknicu o prirodi pogreške. Na koncu, ne smije se zaboraviti i da su istraživanja s analizama pogrešaka pokazala da se vrlo teško određuje što je točno pogreška, koji je točan opseg pogreške, a i gotovo se

svaka pogreška može pripojati različitim kategorijama – a što više kategorija ima, to se kod anotatora stvara dodatna konfuzija i onemogućava mu se donošenje odluka.

Podaci dobiveni analizom pogrešaka iz predistraživanja (Ljubas 2019) omogućuju dijakronijsko praćenje promjena u kvaliteti izlaznih podataka, pogotovo s obzirom na veliku brzinu kojom se razvijaju alati za strojno prevođenje. Promjene u izlaznim podacima gotovo se mogu pratiti iz dana u dan, što znači da odmak od dvije godine između prijevoda generiranog za predistraživanje i prijevoda generiranog za glavno istraživanje predstavlja veliku mogućnost da sustav za strojno prevođenje pokaže napredak u kvaliteti, napose jezičnoj. Pored toga će se dijakronijski popratiti i eventualne promjene u percipiranoj razumljivosti teksta te pomaci u rješavanju testa razumijevanja. Klasificirane će se pogreške kao još jedna važna varijabla dovesti u suodnos s ocjenama percipiranog razumijevanja i brojem točnih odgovora ispitanika.

Svi ovi podaci prikupljeni anketnim obrascem pružit će odgovore na postavljena istraživačka pitanja. Premda se u opisivanju rezultata uključuju i određeni brojčani podaci, u ovom se radu nisu statistički obrađivali ni tako prikazivali jer naglasak nije na analizi podataka dobivenih metodama vrednovanja, koje bi u tom slučaju bile samo instrument provođenja istraživanja, već je fokus na analizi metoda *per se*, njihovih prednosti i mana¹⁵. S jedne bi se strane moglo reći da je takav pristup ograničenje ovog istraživanja, no u razgovorima sa statističarima procijenjeno je da se ne bi sve varijable koje utječu na provođenje metoda mogle obuhvatiti statistikom, da statistička analiza ne bi osobito pridonijela zaključcima o istraživačkim pitanjima te da kvalitativni pristup predstavlja smisleniji pokušaj uspoređivanja i rangiranja metoda. Rezultati istraživanja te rasprava o njima prikazani su u idućem poglavlju.

¹⁵ Pregledom recentne literature Rivera-Trigueros (2022: 613) uočava i druga istraživanja vrednovanja strojnih prijevoda koja se okreću kvalitativnim pristupima te zaključuje da analiza otvorenih odgovora ispitanika ili pak provedenih razgovora s ispitanicima omogućuje pristup informacijama koje istraživaču donose mnogo više znanja i uvida od statističkih analiza odgovora u upitnicima.

5. Rezultati i diskusija

U svakom će se potpoglavlju ovog dijela rada prikazati rezultati kvalitativnog tumačenja podataka dobivenih u istraživanju, prateći u svakom pojedinu istraženu metodu. Na primjerima izlaznih podataka pokazat će se kakve su sve osobitosti obuhvaćenih metoda uočene te će se na temelju kvalitativnog tumačenja podataka i diskusije, gdje to bude moguće, pružiti odgovori na postavljena istraživačka pitanja. Predzadnje potpoglavlje (5.4) posvećeno je dijakronijskoj usporedbi metoda vrednovanja u odnosu na rezultate predistraživanja iz 2019. godine, a u zadnjem potpoglavlju (5.5) predstavlja se opsežnija rasprava o ukupnim zaključcima i saznanjima ovog istraživanja.

5.1 Metoda subjektivne procjene: rezultati i diskusija (IP1, IP2)

U sljedećim potpoglavljima predstavljeni su rezultati subjektivnog ocjenjivanja razumljivosti prijevoda: prvo s fokusom na rezultate koji se odnose na IP1 (*Postiže li se veće slaganje među ispitanicima kada oni ocjenjuju prijevod u cjelini u odnosu na raščlanjene segmente ili obrnuto?*) o kojima je u potpoglavlju 5.1.2 povedena rasprava, a potom s fokusom na rezultate usporedbe ocjena i stavova za IP2 (*Ovise li ocjene pri procjeni razumljivosti teksta o prijašnjim iskustvima sa strojnim prevođenjem ili o stavovima o njemu?*), o kojima je u potpoglavlju 5.1.4 povedena rasprava te kvalitativna analiza stavova ispitanika.

5.1.1 Rezultati vrednovanja subjektivnom procjenom (IP1, rezultati)

U prvoj skupini prosječna ocjena za strojni prijevod, temeljena na aritmetičkoj sredini ocjena svih osamdesetero ispitanika koji su vrednovali raščlanjene segmente, iznosi 4,1. Ta je ocjena za 0,5 viša od prosječne ocjene druge skupine, koja iznosi 3,6, a koju je osamdesetero ispitanika druge skupine dodijelilo razumljivosti čitavog teksta. U Tablici 1 prikazane su sve ocjene koje su ispitanici dodijelili izlaznim podacima, navedene u dvjema skupinama kako su ispitanici i vrednovali tekst prema parametru raščlanjenosti:

Tablica 1: Ocjene razumljivosti

Ispitanik	Aritmetička sredina ocjena raščlanjenih segmenata	Ispitanik	Ocjena cjelokupnog segmenta
I1	2,8	I81	4
I2	4,5	I82	3
I3	4,6	I83	3
I4	4,4	I84	2
I5	4,1	I85	3
I6	3,9	I86	5
I7	3,4	I87	5
I8	4,0	I88	3
I9	3,3	I89	4
I10	3,9	I90	2
I11	4,6	I91	4
I12	4,9	I92	1
I13	4,4	I93	5
I14	4,6	I94	3
I15	4,9	I95	3
I16	4,4	I96	4
I17	3,6	I97	5
I18	3,8	I98	3
I19	3,1	I99	4
I20	3,4	I100	3
I21	4,5	I101	4
I22	4,4	I102	5
I23	4,6	I103	5
I24	4,7	I104	4
I25	3,5	I105	3
I26	4,1	I106	3
I27	4,7	I107	3
I28	3,7	I108	3
I29	4,3	I109	3
I30	4,2	I110	4
I31	3,9	I111	3
I32	4,5	I112	2
I33	4,4	I113	4
I34	4,7	I114	2
I35	4,2	I115	4
I36	4,3	I116	3
I37	4,5	I117	4

I38	3,4	I118	3
I39	4,1	I119	3
I40	3,4	I120	3
I41	4,2	I121	4
I42	3,5	I122	2
I43	4,1	I123	4
I44	3,4	I124	4
I45	3,7	I125	5
I46	3,8	I126	5
I47	5,0	I127	3
I48	4,6	I128	4
I49	4,5	I129	5
I50	4,5	I130	4
I51	3,9	I131	4
I52	3,6	I132	5
I53	3,9	I133	4
I54	4,4	I134	3
I55	3,8	I135	4
I56	4,2	I136	3
I57	4,2	I137	3
I58	4,2	I138	5
I59	4,9	I139	3
I60	5,0	I140	4
I61	3,6	I141	5
I62	5,0	I142	3
I63	3,9	I143	3
I64	3,5	I144	5
I65	3,9	I145	4
I66	4,6	I146	3
I67	4,6	I147	3
I68	5,0	I148	4
I69	4,4	I149	3
I70	3,5	I150	4
I71	4,3	I151	3
I72	3,9	I152	3
I73	3,5	I153	5
I74	3,6	I154	3
I75	4,8	I155	5
I76	4,5	I156	5
I77	3,7	I157	3
I78	4,6	I158	2
I79	3,8	I159	5

I80	3,8	I160	2
Ukupna ocjena:	4,1	Ukupna ocjena:	3,6

Kao što se može iščitati iz Tablice 1 (i kako je postotno prikazano u Grafu 1 u nastavku), od ukupno osamdesetero ispitanika prve skupine:

- četrdeset i sedmero je ispitanika izlazne podatke procijenilo u prosjeku „dobro razumljivima“ (ocjena 4);
- dvadeset i troje je ispitanika izlazne podatke procijenilo u prosjeku „potpuno razumljivima“ (ocjena 5);
- preostalih desetero ispitanika izlazne je podatke procijenilo u prosjeku „prilično razumljivima“ (ocjena 3);
- nijedan ispitanik nije raščlanjene segmente u prosjeku procijenio „slabo razumljivima“ (2) ili „nerazumljivima“ (1).

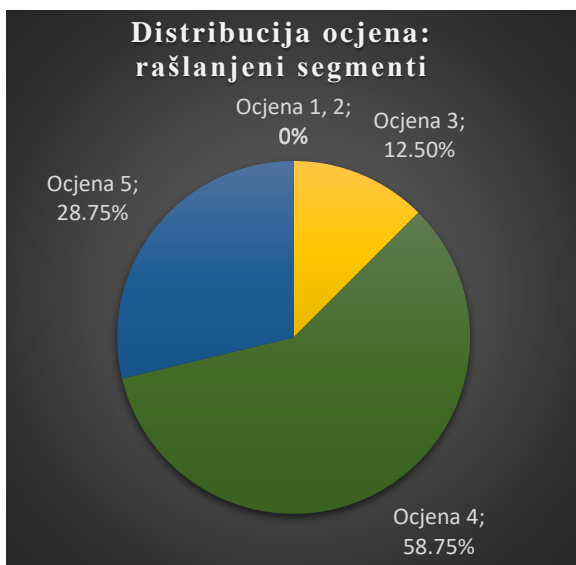
Kako nije uočen velik raspon među dodijeljenim ocjenama, u slučaju ocjenjivanja raščlanjenih segmenata može se govoriti o donekle postignutom slaganju među ispitanicima, premda i dalje relativno niskom.

Skupina koja je izlazne podatke vrednovala kao cjelinu nije ih samo ocijenila nešto lošije, već je postignuta i niža razina slaganja – raspon je ocjena veći i pri odabiru ocjene nije se formirala većina (postotno prikazano u Grafu 2 u nastavku):

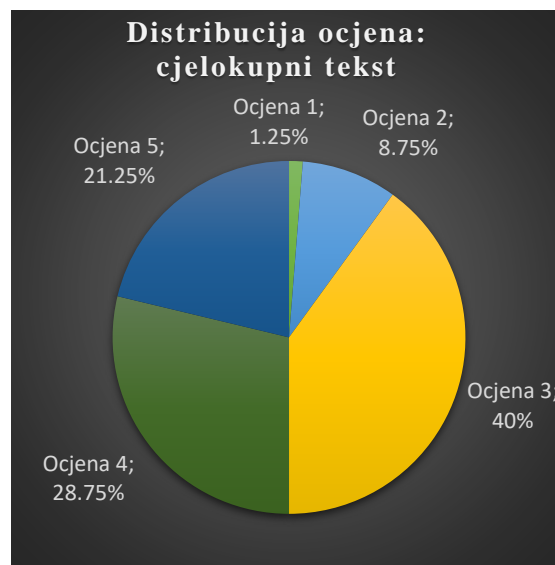
- najveći broj ispitanika, njih trideset i dvoje, izlazne je podatke procijenio „prilično razumljivima“ (ocjena 3);
- dvadeset i troje ispitanika izlazne je podatke procijenilo „dobro razumljivima“ (ocjena 4);
- sedamnaestero ispitanika izlazne je podatke procijenilo „potpuno razumljivima“ (ocjena 5);
- još sedmero ispitanika izlazne je podatke procijenilo „slabo razumljivima“ (ocjena 2);
- jedan ispitanik izlazne je podatke smatrao „nerazumljivima“ (ocjena 1).

U ovoj skupini ne može se govoriti o postignutoj prihvatljivoj razini slaganja.

Ta se distribucija usporedno može promatrati u Grafovima 1 i 2.



Graf 1: Prva skupina ispitanika



Graf 2: Druga skupina ispitanika

U prvoj skupini nema ispitanika koji su prijevodu dali dvije najniže ocjene, a i 58,75 % ispitanika složilo se oko ocjene 4, što pokazuje načelnu tendenciju ispitanika da konkretni prijevod smatraju uspješnim. Ipak, kako se većinsko slaganje od 51 % smatra apsolutnim minimumom slaganja koje je potrebno postići da bi se moglo govoriti i o kakvim relevantnim podacima, ne može se reći da su rezultati osobito vjerodostojni. Odstupanja gdje 28,75 % ispitanika prijevod ocjenjuje ocjenom 5, a 12,5 % ocjenom 3 i dalje su prevelika (v. Graf 1). U svjetlu tog tumačenja jasno je da se 40 % postignutog slaganja za ocjenu 3 u drugoj skupini nikako ne može uzeti u obzir kao relevantan podatak. U drugoj je skupini također 28,75 % ispitanika odabralo ocjenu 4, 21,25 % ocjenu 5, a bilo je i nižih ocjena: 8,75 % ispitanika odabralo je ocjenu 2, i 1,25 % ispitanika ocjenu 1 (v. Graf 2).

5.1.2 Raščlanjeni segmenti ili vrednovanje cjeline? (IP1, diskusija)

Budući da ni u jednoj skupini nije postignuta uvjerljiva razina slaganja među ispitanicima, rezultate ovog dijela istraživanja potrebno je detaljnije kvalitativno proučiti jer se i unatoč neodređenim podacima o slaganju mogu izvući zaključci o vrednovanju cjeline ili raščlanjenih segmenata. Tumačenju širokog dijapazona ocjena i stvarnim razmjerima neslaganja u prvoj skupini pomaže nam analiza vrednovanih segmenata, počevši s Primjerom 1:

Primjer 1. *Ne zbog statusa državnog službenika, koji me ne izlaže neizvjesnostima na tržištu rada.*

Referentni prijevod: *Ne zato što sam zaposlena u državnoj službi, pa zbog toga nisam izložena nepredvidljivosti tržišta rada.*

Izvornik: *Nicht wegen des Beamtenstatus, der mich nicht den Unwägbarkeiten des Arbeitsmarktes aussetzt.*

Riječ je o srednje dugom segmentu koji su ispitanici vrlo različito ocijenili: četrnaestero ispitanika smatra ga „nerazumljivim“ i daje mu najnižu ocjenu 1, a isto toliko ispitanika smatra ga „potpuno razumljivim“ i daje mu najvišu ocjenu 5. Šesnaestero ispitanika ocijenilo ga je „slabo razumljivim“ (ocjena 2), dvadeset i jedan ispitanik „prilično razumljivim“ (ocjena 3) te petnaestero ispitanika „dobro razumljivim“ (ocjena 4). U ovom primjeru do izražaja dolazi najveći nedostatak subjektivne metode procjene – na temelju samih ocjena, koje se toliko razlikuju, istraživač ne može zaključiti ništa o kvaliteti konkretnog prevedenog segmenta. Procjene su posve različite, a podaci ne govore ništa o unutarnjoj logici kojom su se vodili ispitanici kada su se odlučivali za svoje ocjene. Istraživaču preostaje samo da nagađa zašto bi određeni segment nekom ispitaniku bio više ili manje razumljiv.

Što to govori o samoj metodi i načinu provedbe ove metode? Analizirajući Primjer 1. vidimo da je riječ o rečenici čije se puno semantičko značenje i u izvorniku iščitava tek unutar koherentnog teksta. Da bi se doista razumjela, potrebno je poznavati u kojem je odnosu s rečenicama koje joj prethode. I sam referentni prijevod segmenta nije sasvim jasan bez konteksta jer se i on snažno naslanja na prethodne dijelove. Ovaj je segment, dakle, izvlačenjem iz konteksta izgubio mogućnost da prenese puni smisao, čime se može objasniti nejasan ishod subjektivne procjene.

Da bi vrednovanje raščlanjenih segmenata bilo smisljeno, segmenti bi se trebali pisati kontroliranim jezikom (v. potpoglavlje 2.3.3.5 za definiciju kontroliranog jezika) – morale bi se vrednovati samostalne rečenice koje su pročišćene od deiktičkih elemenata, zamjenica ili bilo kojeg drugog elementa koji upućuje rečenicu na kontekst. Stvarni tekstovi, međutim, nisu tako pisani. Oni se eventualno mogu unaprijed pripremiti i preurediti da bi bili tako napisani, ali ako želimo vrednovati razumljivost prirodnih tekstova, onda nam primjeri poput Primjera 1 ukazuju na to da podjela teksta u nasumične rečenične segmente nije dobar izbor jer se ispitanicima izravno time onemogućuje razumijevanje segmenta.

Tomu u prilog govori i Primjer 2, koji je najslabije ocijenjeni segment, prosječnom ocjenom 2,3, iako su kratki segmenti poput ovog u načelu razumljiviji:

Primjer 2. *Pa navijati i kucati u prsa?*

Referentni prijevod: *Znači, imamo se čemu veseliti?*

Izvornik: *Also jubilieren und an die Brust klopfen?*

Iz izvornika i referentnog prijevoda razvidno je da ni ovaj segment ne funkcionira kao samostalna cjelina. Segment je zapravo retoričko pitanje, koje proizlazi iz konteksta koji mu prethodi, a funkcionira kao most prema materijalu koji tek slijedi. Izdvojeno on ne prenosi mnogo značenja. Loša ocjena segmenta stoga se može uzeti samo kao indikator, ali ne i sigurni pokazatelj nerazumljivosti prijevoda, te bi i u ovom primjeru rezultati vrednovanja bili vjerodostojniji da se segment vrednovao u širem kontekstu izlaznih podataka, kao što je navedeno u proširenom Primjeru 3:

Primjer 3. *Zapravo, ne bih se trebao skrivati sa svojim poslom, mogao bih biti ponosan na njega. To je potvrdilo i nedavno istraživanje, prema kojem je 64 posto ispitanih reklo da su učitelji dobro ili čak vrlo dobro radili posao. **Pa navijati i kucati u prsa?** Ali postoji i druga strana medalje. Mislim da teško da postoji zanimanje koje je posljednjih godina tako ozbiljno napadnuto kao učitelji. Postoje razne studije iz Pise, koje više puta dokazuju da su njemački studenti u međunarodnoj usporedbi tek osrednji.*

Referentni prijevod: *Zapravo uopće ne bih trebala skrivati da sam nastavnica, trebala bih moći biti ponosna na to. To mi je nedavno potvrdila i jedna anketa prema kojoj je 64 % ispitanika priznalo da nastavnici dobro ili čak vrlo dobro obavljaju svoj posao. **Znači, imamo se čemu veseliti?** Postoji, međutim, i druga strana medalje. Mislim da nema te struke koja je proteklih godina bila tako snažno napadnuta kao nastavnička. Iskaču neke PISA-ine studije koje neprestano dokazuju da su naši učenici u međunarodnoj usporedbi tek osrednji.*

Izvornik: *Eigentlich müsste ich mich mit meinem Beruf nicht verstecken, könnte stolz auf ihn sein. Das bestätigte mir vor Kurzem eine Umfrage, nach der 64 Prozent der Befragten den Lehrern bescheinigten, dass sie gute oder sogar sehr*

gute Arbeit leisteten. Also jublieren und an die Brust klopfen? Da ist aber auch die andere Seite der Medaille. Ich glaube, kaum ein Berufsstand wurde in den vergangenen Jahren so stark angegriffen wie der der Lehrer. Da sind die diversen Pisa-Studien, die immer wieder belegen, dass deutsche Schüler im internationalen Vergleich gerade mal nur Mittelmaß sind.

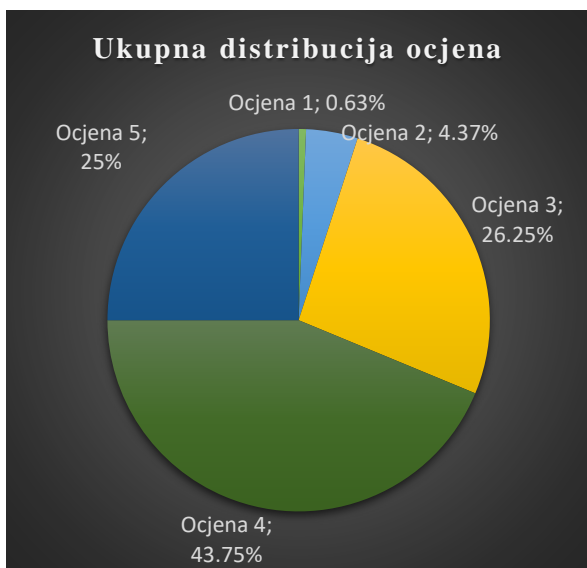
Sa širim pogledom na tekst ispitanik lakše može odrediti zašto je došlo do određenih pogrešaka u prijevodu ili može i unatoč pogreškama dokučiti generalni smisao odlomka, čime se neutralizira važnost pogreške u sitnijim elementima.

Prvi je zadatak postavljen ne bi li se vidjelo može li se postići veće slaganje među ispitanicima ako vrednuju raščlanjene segmente ili cjelokupni tekst. Premda je blago veće slaganje postignuto među ispitanicima prve skupine koji su vrednovali raščlanjene segmente, zaključeno je da to slaganje nije ni približno dovoljno visoko da bi podaci bili relevantni, pa – samo gledano iz kuta razine slaganja – i nije moguće odgovoriti na IP1 i dati jasnu prednost jednom ili drugom načinu vrednovanja. S druge strane, kvalitativnom analizom primjera uočeno je da raščlanjivanje segmenata čak i u izvorniku ili referentnom prijevodu već utječe na otežano razumijevanje. Ti su nalazi u skladu s recentnijim tendencijama reevaluacije ustaljenih praksi u vrednovanju strojnog prevođenja. Kao što su, primjerice, Läubli i sur. u svom istraživanju iz 2020. pokazali, i ovo istraživanje potvrđuje da nema jasnih razloga zašto bi se sa sve boljim izlaznim podacima pri provedbi metode subjektivne procjene zadržala starija praksa evaluacije tekstova na principu segmenata raščlanjenih na rečenice, jednako kao što se strojnih prijevodi ne bi mogli vrednovati na razini riječi ili drugih manjih elemenata. Ovo je istraživanje pokazalo da ispitanici smislenije vrednuju tekstove u cjelini, kada u obzir mogu uzeti punu koheziju i koherentnost: tekst je tekst samo kada je cjelovit, a ni ljudski se prijevodi ne vrednuju raščlanjeno. Upravo zato i pri vrednovanju strojnih prijevoda prednost valja dati vrednovanju cjeline.

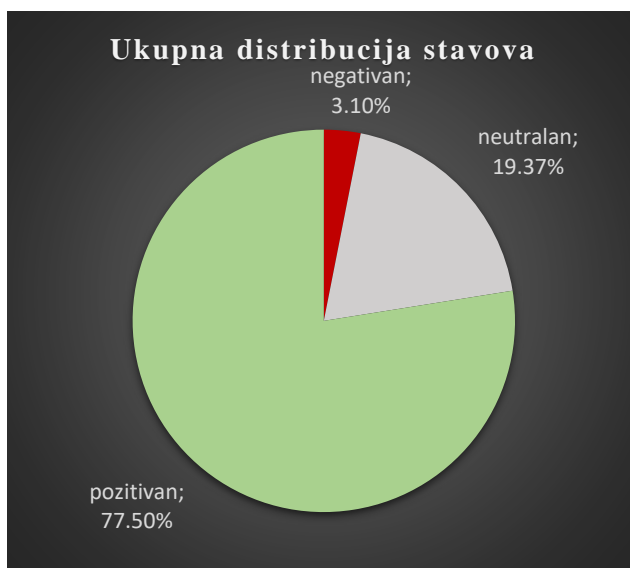
5.1.3 Suodnos ocjena i prijašnjih stavova ispitanika (IP2, rezultati)

Kao što je prikazano u teorijskom dijelu, starija su istraživanja strojnog prevođenja stalno iznova bilježila da su ispitanici općenito negativno nastrojeni prema strojnom prevođenju, što se odražavalo i na subjektivnu procjenu kvalitete izlaznih podataka. Danas u 2023. godini

možemo govoriti o popriličnom pomaku u tom smislu: bilo da se razlog krije u drastičnom poboljšanju kvalitete izlaznih podataka ili u pozitivnoj reinterpetaciji ideje strojnog prevođenja među općom populacijom, samo je jedan od svih stotinu šezdeset ispitanika ovog dijela istraživanja izlazne podatke procijenio „nerazumljivima“, a samo je još sedmero ispitanika prijevod ocijenilo „slabo razumljivim“. Vidljivo je iz Grafa 3 da to čini 5 % ispitanika koji su prijevodu dali najniže ocjene 1 i 2:



Graf 3: Obje skupine ispitanika



Graf 4: Stavovi ispitanika

Srednje razumljivima izlazne su podatke procijenili četrdeset i dvoje ispitanika odnosno njih 26,25 %, a više od dvije trećine, 68,75 % ispitanika, prijevodu je dalo visoke ocjene: njih sedamdesetero procijenilo ga je „dobro razumljivim“, a četrdesetero „potpuno razumljivim“.

S ukupnom distribucijom ocjena na Grafu 3 može se usporediti i distribucija stavova na Grafu 4 koji se odnosi na stavove ispitanika čiji će se utjecaj na ocjene analizirati u sljedećem potpoglavlju. Ovdje se donose rezultati anketnog obrasca o stavovima i iskustvima ispitanika (v. Prilog C): uočeno je da pretežito negativan stav prema strojnom prevođenju ima tek petero ispitanika (3,1 %), neutralan stav ima trideset i jedan ispitanik (19,37 %), a pretežito pozitivan stav ima više od dvije trećine ispitanika, njih čak sto dvadeset i četvero (77,5 %). Devedeset i osmero ispitanika potvrdilo je da se u svom profesionalnom životu (a koji se ne odnosi na prevoditeljsku profesiju) kao i u privatnom životu aktivno služi alatima za strojno prevođenje, a sto trinaest ispitanika potvrdilo je da su automatski prijevodi u njihovim životima prisutni, primjerice, na društvenim mrežama ili kada čitaju mrežne stranice. To svjedoči o širokoj

rasprostranjenosti alata za strojno prevođenje danas. Načelno pozitivan stav ispitanika također se može rastumačiti stupnjem tolerancije prema pogreškama u strojno prevedenim tekstovima koji raste s učestalošću izloženosti ispitanika strojno prevedenim tekstovima.

Čak sto četrdeset i jedan ispitanik, od njih ukupno sto šezdeset (88,1 %), smatra da su alati za strojno prevođenje ljudima korisni u svakodnevnim situacijama. Nešto umjereniji broj ispitanika, njih sedamdeset i troje (45 %) pozitivno je nastrojeno prema ideji razvitka strojnog prevođenja i ne smatraju da alati za strojno prevođenje otimaju posao stručnim prevoditeljima, što se u prevoditeljskim krugovima katkad čuje kao argument protiv njihove sve šire uporabe, a s čime se slaže ili donekle slaže trideset i jedan ispitanik (19,3 %). Sto sedamnaestero ispitanika (73,1 %) već je zadovoljno i trenutnim stanjem izlaznih podataka na hrvatskom jeziku za njihove potrebe te su optimistični i za daljnji razvitak alata za strojno prevođenje.

5.1.4 Analiza stavova ispitanika (IP2, diskusija)

Formuliranjem IP2 (*Ovise li ocjene pri procjeni razumljivosti teksta o prijašnjim iskustvima sa strojnim prevođenjem ili o stavovima o njemu?*) prvenstveno se htjelo proučiti utječu li negativni stavovi ispitanika ujedno na to da ispitanici negativno ocjenjuju izlazne podatke, što je tendencija zamijećena u starijim istraživanjima strojnog prevođenja. U nastavku se izlaže zašto ovo istraživanje ne potvrđuje da je to i dalje slučaj. Taj je zaključak izveden analizom razina slaganja svih sto šezdeset ispitanika na Likertovoj ljestvici s nizom pozitivnih i negativno izraženih izjava (v. Prilog C), no u nastavku je diskusija upotpunjena i dijelom odgovora slobodnog tipa u slučajevima kada su se ispitanici poslužili tom mogućnošću da bi dodatno pojasnili svoje stavove i time na slobodniji i personalniji način pružili uvid u svoje potrebe i očekivanja od strojnog prevođenja.

Od petero ispitanika koji su pokazali negativne stavove prema strojnom prevođenju, ovdje izdvajamo tri slučaja. Ispitanik koji je izrazio najnegativniji stav prema strojnom prevođenju predmetni je strojni prijevod procijenio „potpuno razumljivim“ i dao mu ocjenu 5. I drugi ispitanik kod kojega su zabilježeni negativni stavovi prema strojnom prevođenju tekstu je dodijelio ocjenu 4,1, koja je također visoka, a odgovara i prosječnoj ocjeni razumljivosti. Ovaj se ispitanik (I5) od prvog razlikuje u tome što se u svom profesionalnom životu često susreće s

alatima za strojno prevođenje. U dijelu ankete gdje je bilo moguće slobodno upisati svoje mišljenje o alatima za strojno prevođenje, taj ispitanik piše sljedeće¹⁶:

I5, pravnik: Njihovu korisnost vidim ponajprije i ponajviše u pronalaženju prijevoda za pojedine pojmove, potrazi za sinonimima i sl., a manje u slučaju potrebe za prijevodom rečenica (pogotovo dužih i složenijih), odlomaka ili cijelih tekstova. (...) Što se tiče razumljivosti strojnih prijevoda, čini mi se da je konkretno Google Translate doživio znatan napredak u kvaliteti i razumljivosti prijevoda (čak i potpunih rečenica ili odlomaka) u zadnjih 5-10 godina.

Iz ovog se stava može iščitati na koje se sve načine neprevoditelji u profesionalnom životu ipak služe alatima za strojno prevođenje. Ispitanik sam izdvaja Google Prevoditelj, premda su ispitanici iz uputa samo znali da rade sa strojnim prijevodom, ali ne i kojim je alatom nastao. Donekle čudi što se I5 opredijelio kao negativno nastrojen prema alatima za strojno prevođenje te što ne vjeruje u njihov daljnji napredak, pogotovo uzevši u obzir da je već i sam zamijetio promjene nabolje u Google Prevoditelju te ostavlja dojam kao da o tome aktivno promišlja, a usto je i tekst istraživanja ocijenio „dobro razumljivim“. Razlozi za to mogu biti mnogi i individualni: neki ispitanici možda općenito imaju strah od automatiziranja ljudskog rada, ili odgovaraju na pitanja onako kako smatraju da bi bilo poželjno, ili pak razlozi nisu strogo povezani sa samim predmetom istraživanja. Kod ovakvih *black box* metoda vrednovanja, istraživač o motivaciji odgovora ispitanika može samo nagađati.

Samo treći ispitanik (I92) s uočenim negativnim stavovima svoje je stavove popratio i negativnom ocjenom za predmetni prijevod. Strojno prevođenje komentira na sljedeći način:

I92: Strojno prevođenje nije baš najbolje. Ono je super za prevesti frazu ili riječ, ali ne i cijelu rečenicu jer npr. Google Translate će doslovce svaku riječ u rečenici prevesti te ona gubi svoj kontekst i smisao.

I ovaj ispitanik izdvaja Google Prevoditelj te potvrđuje da ima i pozitivna iskustva s prijevodom kraćih tekstualnih elemenata, ali izražava nezadovoljstvo prijevodima čitavih rečenica, što je u

¹⁶ Komentari ispitanika preuzeti su kao direktni, neuređeni i nelektorirani citati, osim kada je znakom (...) naznačeno kraćenje dugačkih komentara.

skladu s njegovim negativnim stavom. No, kao što je već navedeno, ispitanici negativnih stavova u ovom su istraživanju bili u manjini.

S druge strane, ispitanici kod kojih su zabilježeni pozitivni stavovi prema strojnom prevođenju nisu pokazali tendenciju da zbog toga automatski prijevod ocjenjuju potpuno razumljivim. Šestero je ispitanika pretežno pozitivnih stavova prema strojnom prevođenju konkretni prijevod ocijenilo „slabo razumljivim“ (ocjena 2), a dvadeset i devetero samo „prilično razumljivim“ (ocjena 3). Ispitanik broj 116, koji ovaj prijevod procjenjuje prilično razumljivim, strojnim se prevođenjem služi u svom profesionalnom životu, i komentira najbolje strane alata za strojno prevođenje:

I116, psiholog: *Nisu savršeni, ali iznimno korisni, najbolje je što su tako brzi i praktični.*

U komentarima ispitanika čije se ocjene razumljivosti poklapaju s generalnim pozitivnim stavovima prema alatima za strojno prevođenje također se zamjećuje da i oni, kao prevoditeljski ili lingvistički laici, svejedno razmišljaju o prijevodima koji su im javno dostupni te njihovom dijakronijskom razvoju, kao što to čini I48, koji je ovom prijevodu dodijelio ukupnu ocjenu 4,6:

I48, marketinški stručnjak: *Mislim da je prijevod dosta dobar, Čak mi se čini bolji nego prije nekih 5 god, ponekad je teže kada se radi o stručnijim riječima i kontekstu, ali nije mi to baš problem.*

Na koncu, iz komentara ispitanika koji su pozitivno nastrojeni prema strojnom prevođenju te su i ovaj testni prijevod smatrali „potpuno razumljivim“ (ocjena 5) razvidno je da oni od alata za strojno prevođenje ne očekuju savršene prijevode na razini konvencionalnih prijevoda, već da njihov pozitivan stav proizlazi iz toga što realno sagledavaju mogućnosti alata i što se njima služe u situacijama u kojima su im doista korisni, napose za brzinsko razumijevanje kraćih tekstualnih elemenata ili shvaćanja generalne poruke teksta, u poslovnom, akademskom ili privatnom okruženju:

I37, magistar znanosti struke o okolišu: *Vrlo su korisni za prijevod kraćih fraza ili shvaćanje konteksta dužih komada teksta, ali ne mogu prenijeti sve 'finese'.*

I93, psiholog: *Meni je vrlo korisno u situacijama kada mi brzo treba prijevod jezika kojeg uopće ne poznajem. Osim toga, korisno mi je i za čitanje stručne literature kada samo želim 'uloviti glavnu misao'.*

Može se zaključiti da je kod velike većine ispitanika različitih struka i razina obrazovanja zamijećeno da su već došli u doticaj s alatima za strojno prevođenje i da se njima služe u svrhe u kojima ih smatraju prikladnima i korisnima. U prosjeku su se stavovi neprevoditelja prema strojnom prevođenju u Hrvatskoj pokazali pozitivnima. Kao odgovor na istraživačko pitanje može se istaknuti da nije utvrđeno da ispitanici svjesno ili nesvjesno upravljaju svojim subjektivnim procjenama kako bi alate za strojno prevođene prikazali u negativnijem ili pozitivnijem svjetlu. Zaključak je ovog dijela istraživanja da ocjene pri subjektivnoj procjeni razumljivosti danas u manjoj mjeri ovise o općenitim stavovima ispitanika prema strojnom prevođenju nego što je to bio slučaj u prijašnjim istraživanjima.

Premda odgovor na IP2 načelno govori u prilog metodi subjektivnog vrednovanja, ipak još ostaje razmotriti koliko su rezultati te metode generalno relevantni, s obzirom na to da se teško uočavaju obrasci zašto određeni ispitanici podacima daju određene ocjene te da se ni promjenama varijable raščlanjenosti nije postiglo relevantno slaganje u procjenama ispitanika. O pitanju vjerodostojnosti podataka dobivenih metodom subjektivne procjene rasprava će se nastaviti u potpoglavlju 5.5, nakon pregleda rezultata svih metoda vrednovanja.

5.2 Metoda testa razumijevanja: rezultati i diskusija (IP3)

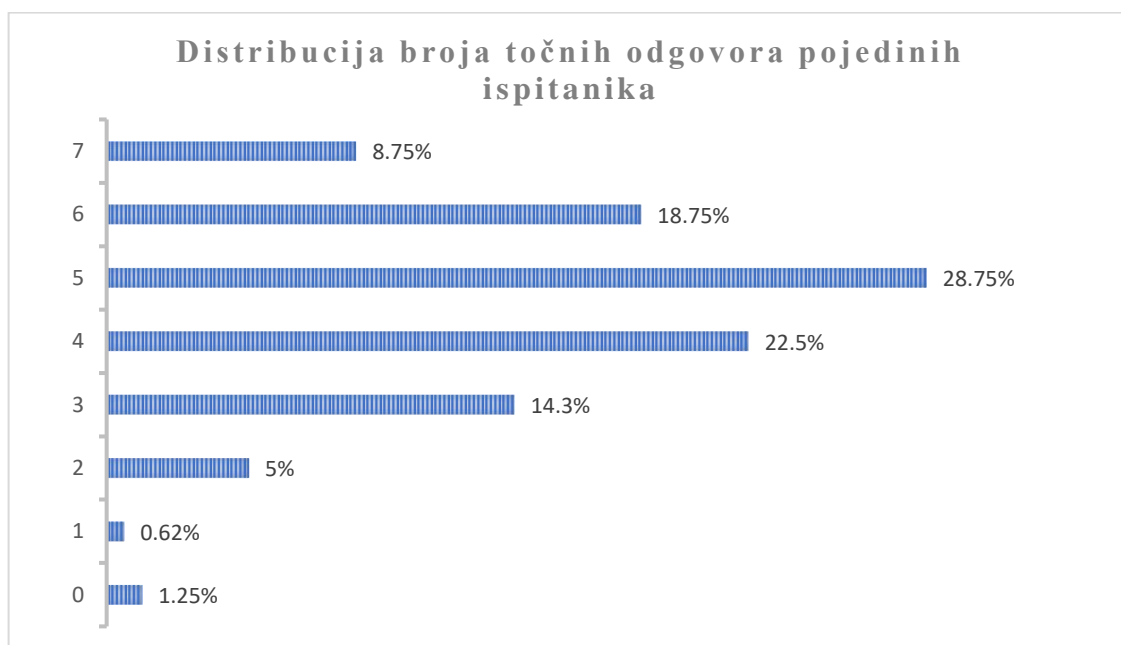
Metodom testa razumijevanja pokušava se doskočiti subjektivnosti metode procjene razumljivosti. Rezultati te metode objektivniji su već time što broj točnih odgovora koju su ispitanici ponudili na testu daje jasan uvid u kvalitetu prijevoda određenog teksta ili dijela teksta. U sljedećim potpoglavljima prikazani su rezultati metode testa razumijevanja u ovom istraživanju, a zatim se analizira suodnos broja točnih odgovora na testu s kvalitetom prijevoda, a potom i sa subjektivnim ocjenama razumljivosti kako bi se pružio odgovor na IP3 (*Jesu li ocjene pri procjeni razumljivosti teksta proporcionalne broju točnih odgovora na testu razumijevanja?*).

5.2.1 Broj točnih odgovora na testu razumijevanja

U Prilogu B prikazana je distribucija točnih i netočnih odgovora svih ispitanika na sva pitanja testa iz koje je razvidno da:

- dvoje ispitanika nije točno odgovorilo ni na jedno pitanje;
- jedan je ispitanik točno odgovorio samo na jedno pitanje;
- osmero ispitanika točno je odgovorilo na dva pitanja;
- dvadeset i troje ispitanika točno je odgovorilo na tri pitanja;
- trideset i šestero ispitanika točno je odgovorilo na četiri pitanja;
- četrdeset i šestero (najveći broj ispitanika) točno je odgovorilo na pet pitanja;
- tridesetero ispitanika točno je odgovorilo na šest pitanja (samo jedan pogrešan odgovor);
- četrnaestero ispitanika točno je odgovorilo na sva pitanja u ponuđenom testu razumijevanja.

U Grafu 5 u postocima je prikazana ova distribucija:



Graf 5: Distribucija broja točnih odgovora – svi ispitanici

Iz Grafa 5 razvidno je da je gotovo 80 % ispitanika točno odgovorilo na više od pola pitanja: od toga najviši postotak ispitanika, 28,75 %, ima pet točnih odgovora, a 22,5 % ima četiri točna

i 18,75 % ima šest točnih odgovora. Zanimljivo nije ni postotak od 8,75 % ispitanika koji su točno odgovorili na sva postavljena pitanja. Taj je podatak već opipljiv dokaz o kvaliteti izlaznih podataka, a kako bi se dodatno provjerilo jesu li ti podaci objektivni, ispitanici su na ljestvici od 1 do 5 odredili koliko su sigurni u svoje odgovore (v. Prilog E za usporedbu ocjena subjektivnosti, broja točnih odgovora te sigurnosti u odabrane odgovore), a također postoji podatak o tome koliko su vremena uložili u rješavanje testa.

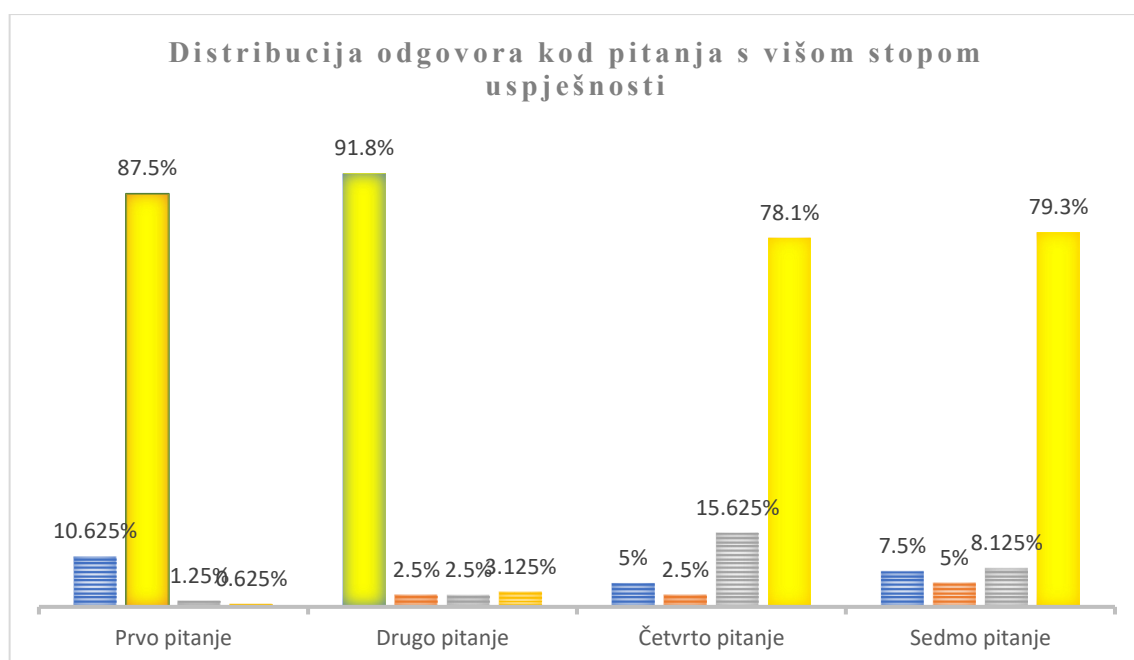
Iz analize tih podataka vidi se da je jedan od dvojice ispitanika (I57) koji nisu odgovorili točno ni na jedno pitanje čitav test riješio za samo jednu minutu, a i sam je priznao da nimalo nije siguran u svoje odgovore. Ako ispitanik nije uložio nimalo truda u potrazi za točnim odgovorima, logično je da će se njegovi rezultati isticati kao loši, i tada oni više govore o nemotivaciji dotičnog ispitanika za sudjelovanje u istraživanju, nego doista o razumljivosti ili nerazumljivosti teksta. Upravo je zbog ovakvih slučajeva kod metode testa razumijevanja važno provjeriti koliko su truda ispitanici uložili u rješavanje testa.

Iz tih je podataka uočeno da je kod drugog ispitanika koji nije točno odgovorio ni na jedno postavljeno pitanje posrijedi druga situacija. Ovaj je ispitanik (I74) na rješavanje testa uložio dvadeset minuta i osjećao se sasvim sigurnim u svoj odabir odgovora, premda su oni svi bili netočni. Pritom je riječ o ispitaniku s pozitivnim stavovima prema strojnom prijevodu koji je tekst procijenio „dobro razumljivim“ (ocjena 4). Dok se takve situacije javljaju na bazi iznimke, a ne statistički relevantnih fenomena, treba ih prihvatiti kao prirodne posljedice ljudskog faktora u istraživanjima. Rješavanje testova razumijevanja je, naime, određena vještina za sebe. Neki su se ljudi tijekom života češće našli u situaciji da sudjeluju u takvim vrstama provjere (jezičnih) znanja, a drugi rjeđe, i to se može odraziti na njihovu uspješnost u rješavanju i ovog testa.

S druge strane, imamo primjer ispitanika (I17) koji je pružio tek dva točna odgovora, ali je i sam izrazio bojazan da je odabrao većinu netočnih odgovora. Njegova svijest o tome da vjerojatno nije dobro riješio test razumijevanja pokazuje da je ispitanik doista imao znatnih teškoća s razumijevanjem izlaznih podataka i njegovo iskustvo svakako ne treba odbaciti. No kako je i metoda testa razumijevanja jedna od metoda kojom se relativno lako može prikupiti veći broj ispitanika (za razliku od npr. metode analize pogrešaka), tako se jasno može vidjeti je li to nešto što se ponavlja kod većeg broja ispitanika ili, kao u ovom slučaju, ipak ostaje na razini iznimke.

5.2.2 Suodnos broja točnih odgovora s kvalitetom strojnog prijevoda

Metoda testa razumijevanja pruža i uvide u to koji su dijelovi teksta bolje prevedeni od drugih. Neka su pitanja (prvo, drugo, četvrto i sedmo, v. Prilog B) ispitanicima predstavljala manje teškoće, odnosno većina je na njih odgovorila točno, dok su na preostalim pitanjima (trećim, petim i šestim, v. Prilog B) u mnogih slučajevima te teškoće bile nepremostive. Graf 6 prikazuje postotnu uspješnost ispitanika na četiri pitanja koja su zadovoljavajuće riješili, dok Graf 7 (v. str. 101) prikazuje postotnu uspješnost ispitanika na lošije riješenim pitanjima. Stupci s točnim odgovorima označeni su žutom bojom kako bi se jasno vidjelo koliko su uspješno ispitanici pokazali razumljivost kod ovih pitanja.



Graf 6: Postotna uspješnost ispitanika na prvom, drugom, četvrtom i sedmom pitanju

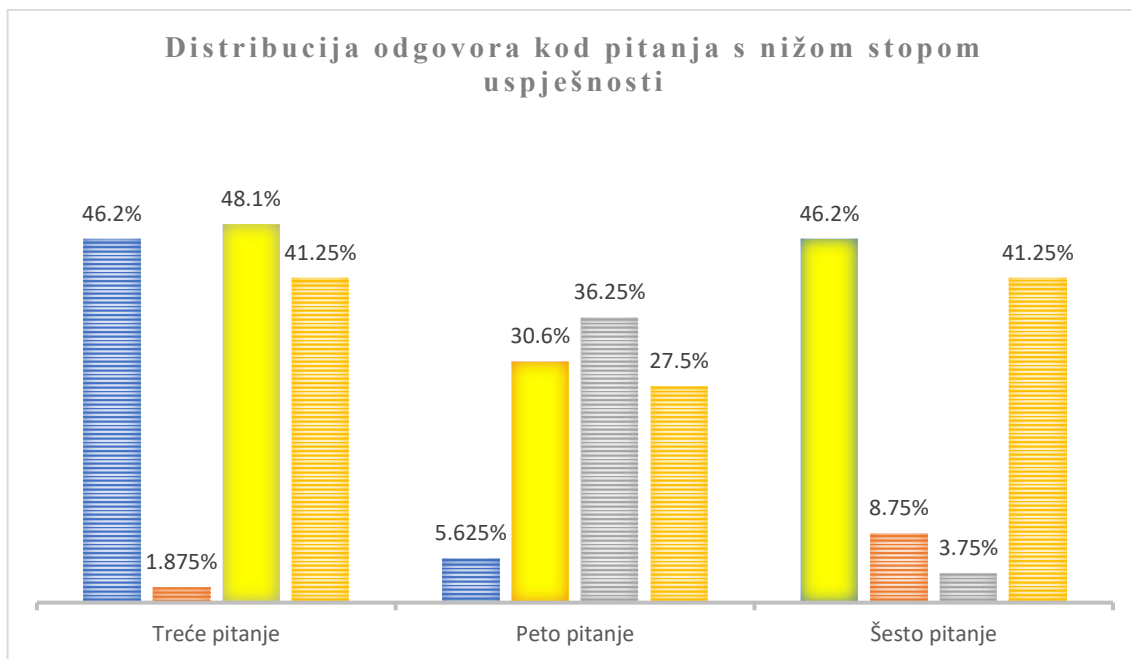
Tako je na drugo pitanje („Zašto se osoba odlučila za nastavničko zanimanje?“) čak sto četrdeset i sedmero ispitanika, odnosno njih 91,8 % odabralo točan odgovor A: *Htjela je mlade podučavati stvarima važnima za život*. Četvero je ispitanika odabralo odgovor B: *Tražila je posao koji joj pruža mnogo slobode* i odgovor D: *Tražila je odgovoran posao*; petero ispitanika odabralo je odgovor C: *Sviđala joj se sigurnost tog radnog mjesta*. Kako su svi potencijalni odgovori na pitanje koncipirani tako da su vjerojatni i da odgovor na ovo konkretno pitanje doista treba pronaći u tekstu, ispitivač iz visokog broja točnih odgovora s relativnom sigurnošću

može zaključiti da je drugi odlomak teksta u kojemu se krio odgovor na ovo pitanje razumljivo preveden.

Ispitanici su u velikom broju točno odgovorili i na prvo pitanje („Kako se osoba osjeća kada govori o svom zanimanju?“): njih sto četrdesetero odnosno 87,5 % odabralo je točan odgovor B: *Ima pomiješane osjećaje*. Ispitanici koji su na pitanje odgovorili netočno, najčešće su odabirali odgovor A: *Osjeća se vrlo važno*, koji je pogrešno odabralo sedamnaestero ispitanika, dok je samo dvoje odabralo odgovor C: *Ljuti se na one koji je to pitaju* i samo jedan ispitanik odgovor D: *Ne osjeća ništa*. Pritom je ovdje riječ o pitanju čiji se odgovor nije mogao naći samo na jednom točnom mjestu, primjerice u jednoj rečenici, već je trebalo pročitati nekoliko odlomaka teksta i smisljeno povezati informacije, što dodatno svjedoči o razumljivosti teksta.

Slična se situacija javlja i kod odgovora na četvrto pitanje („S obzirom na opisanu situaciju, ova se osoba zalaže za to da...“) i na sedmo pitanje („Prva iskustva s vrednovanjem nastavnika pokazuju da...“) na koja je sto dvadeset i petero (78,1 %) odnosno sto dvadeset i sedmero (79,3 %) ispitanika odabralo točne odgovore D (*...nastavnici razmisle o tome kako sami mogu promijeniti neke stvari nabolje; odnosno ...tek rijetki nastavnici razgovaraju o kritici koju su dobili od učenika*). Na četvrto pitanje osmero je ispitanika odgovorilo A: *...roditelji preuzmu veću odgovornost za odgoj svoje djece*, a samo dvoje odgovor C: *...se učenicima smanji pritisak i opterećenost uspjesima*. Nešto više ispitanika, njih dvadeset i petero, dalo se navesti na odgovor B: *...društvo pruži bolju podršku nastavnicima*, no i ovdje je riječ o pitanju za koje ne postoji eksplicitno izrečeni odgovor u tekstu, već ga je potrebno iščitati iz indirektnih izjava. Na sedmo pitanje dvanaestero je ispitanika odabralo pogrešan odgovor A: *...su tek rijetki učenici spremni pružiti konstruktivnu kritiku*; osmero ispitanika odgovor B: *...učenici brzo izgube volju za sudjelovanjem*; i trinaestero ispitanika odgovor C: *...se učenici boje da će se ugroziti njihovo pravo na anonimnost*.

Na preostala tri pitanja manje je od polovice ispitanika odgovorilo točno, što je razvidno iz Grafa 7. Stupci s točnim odgovorima ponovno su označeni žutom bojom, a iz grafa je već na prvi pogled vidljivo da točni odgovori nisu u svim pitanjima prednjačili.



Graf 7: Postotna uspješnost ispitanika na trećem, petom i šestom pitanju

Na treće pitanje („Kakav je, prema mišljenju ove osobe, stav javnosti prema nastavničkom zanimanju?“) sedamdeset i sedmero ispitanika (48,1 %) odgovorilo točan odgovor C: *Nastavnici su istovremeno i vrlo priznati i snažno ih se kritizira*, dakle nešto manje od polovice ispitanika. Ispitanici koji su pogrešno odgovorili, najčešće su odabirali odgovor A: *Nastavnike se krivi za sve što je u društvu loše* (46,2 %) i odgovor D: *Nastavnike se uglavnom smatra nesposobnima* (41,25 %), a tek je troje ispitanika (1,87 %) odgovorilo B: *Negativan stav prema nastavnicima više nije toliko izražen kao prije*. Odgovori A i D prenose sličnu poantu, i to onu koja je djelomično sadržana i u točnom odgovoru. Naime, u kolumni se doista raspravlja o negativnom viđenju javnosti prema nastavnicima, pa je pretpostavka da su ispitanici koji su odabrali ta dva odgovora vidjeli mjesta u tekstu koja to tematiziraju i pomislili su da su time pronašli odgovor. Na pogrešan trag mogla ih je, primjerice, navesti rečenica:

Primjer 4. *Mislim da teško da postoji zanimanje koje je posljednjih godina tako ozbiljno napadnuto kao učitelji.*

Referentni prijevod: *Mislim da nema te struke koja je proteklih godina bila tako snažno napadnuta kao nastavnička.*

Izvornik: *Ich glaube, kaum ein Berufsstand wurde in den vergangenen Jahren so stark angegriffen wie der der Lehrer.*

Ova rečenica pritom nije dovoljna – trebalo je povezati informacije i iz drugih dijelova teksta, kada se govorilo i o pozitivnim stavovima javnosti prema nastavnicima.

Na šesto pitanje („Osoba smatra da bi bilo dobro kada bi...“) točan je odgovor A: *...nastavnici zajedničkim snagama osvježavali didaktičke pristupe u nastavi* odabralo sedamdeset i četvero ispitanika (46,2 %), ali je i šezdeset i šestero ispitanika (41,25 %) odabralo netočan odgovor D: *...se u suradnji s učenicima razvijali novi modeli podučavanja i učenja*. Odgovor na ovo pitanje krio se u jednoj riječi, što se vidi u Primjeru 5:

Primjer 5. *I što mi poduzimamo po tom pitanju? Pripremaju li učitelji predmetne nastave zajedno? Pokušavamo li alternativne metode? Razvijamo li zajednički oblike komunikacije koji nas približavaju željama i idejama naših učenika? Nažalost ne. U većini slučajeva svaki kolega ostaje tradicionalni usamljeni ratnik, a ono što se događa iza vrata razreda ostaje tajna.*

Referentni prijevod: *A kako se borimo protiv toga? Pripremaju li predmetni nastavnici zajednički lekcije? Iskušavamo li alternativne metode? Razvijamo li zajednički oblike komunikacije koji bi nas približili željama i predodžbama naših učenika? Nažalost, ne. U većini slučajeva svaki kolega bori se sam za sebe, a ono što se odvija iza zatvorenih vrata učionice ostaje tajna.*

Izvornik: *Und was tun wir dagegen? Bereiten Fachlehrer gemeinsam Unterrichtseinheiten vor? Probieren wir alternative Methoden aus? Entwickeln wir gemeinsam Kommunikationsformen, die uns näher an die Wünsche und Vorstellungen unserer Schüler bringen? Leider nicht. In der Mehrzahl der Fälle bleibt jeder Kollege der traditionelle Einzelkämpfer und das, was sich hinter der Klassentür abspielt, bleibt geheim.*

Autorica kolumne postavlja (retorička) pitanja o međusobnoj suradnji nastavnika, a na koncu na njih i odgovara, izražavajući žaljenje zbog toga što suradnja nastavnika izostaje. U riječi „nažalost“ krije se objašnjenje zašto je odgovor A točan. To što je toliko ispitanika odabralo odgovor D može biti povezano s time što je zadnji odlomak teksta u potpunosti posvećen vrednovanju učenika koje su neke škole uvele, pa je to ispitanike moglo navesti da misle kako je upravo to vrednovanje jedan od modela koji se ostvaruje u suradnji s učenicima. Odgovor B: *...nastavnici snažno usmjerili pažnju javnosti na svoje probleme* koji je odabralo četrnaestero

ispitanika (8,75 %) u principu se ne tematizira u tekstu, osim što se javnost spominje na početku u kontekstu kritiziranja nastavnika, a odgovor C: *...kolege vrednovali rad svakog nastavnika* koji je odabralo tek šestero ispitanika (3,75 %) pokušava zbuniti ispitanike činjenicom da se u tekstu doista spominju načini vrednovanja nastavnika, samo što ih neće vrednovati kolege, već učenici. Šesto je pitanje vrlo dobar pokazatelj robusnosti pitanja – svaki od ponuđenih odgovora u stvarnosti bi mogao doći u obzir kao odgovor na pitanje, i ako je koji dio teksta nerazumljivije preveden, to lako može zavarati ispitanike. Katkad je i jedna pogreška dovoljna, kada je riječ o nekom ključnom konceptu, da se ispitanici nađu na pogrešnom tragu.

Dobro se to vidi u petom pitanju. Na peto pitanje („Osoba smatra da se nastavnici boje...“) devetero je ispitanika (5,62 %) odgovorilo A: *...da nisu u tijeku s najnovijim znanstvenim i stručnim spoznajama*; a ostali su se ispitanici manje-više ravnomjerno raspodijelili na preostala tri odgovora. Dapače, najveći broj ispitanika, njih pedeset i osmero (36,25 %) odabralo je pogrešan odgovor C: *...razgovarati s kolegama o svojim problemima*, a tek onda slijedi točan odgovor B: *...da ih učenici neće dobro prihvatiti* koji je odabralo četrdeset i devetero ispitanika (30,6 %). Također je odgovor D: *...u nastavi iskušati nešto novo* odabralo četrdeset i četvero ispitanika (27,5 %). Iz toga je razvidno da je ovaj dio teksta nerazumljiviji i nekvalitetnije preveden. Usto, točan odgovor nije eksplicitno izražen, već se krije u dubinskom razumijevaju sljedećeg odlomka:

Primjer 6. *Ako mi učitelji stavljamo karte na stol jedni drugima bez sramežljivosti i lažnog srama, svi moramo priznati da je naša svakodnevica popraćena trajnim strahom - strahom od neuspjeha pred razredom. Moramo priznati da nijedna lekcija nije čista rutina, već upravo suprotno: uvijek iznova borba za priznanje, borba za pažnju i disciplinu.*

Referentni prijevod: *Kada bismo si mi nastavnici bez straha i lažnog srama položili karte na stol, morali bismo priznati da nam je svakodnevica popraćena neprekidnim strahom – strahom od neuspjeha pred razredom. Morali bismo priznati da nijedan nastavni sat nije čista rutina, već upravo suprotno: uvijek je on iznova borba za priznanjem, pažnjom i disciplinom.*

Izvornik: *Wenn wir Lehrer untereinander ohne Scheu und falsche Scham die Karten auf den Tisch legen, müssen wir alle zugeben, dass unser Alltag von permanenter Angst begleitet wird – der Angst, vor der Klasse zu scheitern. Wir*

müssen zugeben, dass keine Unterrichtsstunde reine Routine ist, sondern das Gegenteil: immer wieder ein Kampf um Anerkennung, ein Ringen um Aufmerksamkeit und Disziplin.

U ovom se odlomku javlja pogreška koja bi mogla objasniti zašto je toliko ispitanika odabralo netočan odgovor C. Njemački prilog *untereinander* u ovoj rečenici znači *među sobom* i odnosi se na činjenicu da si nastavnici, kao članovi iste struke, među sobom, među sebi sličnima (a ne pred učenicima, roditeljima, ravnateljima ili drugima koji nisu dio njihove struke) mogu priznati da se boje neuspjeha pred razredom, pred učenicima. U izlaznim podacima odabrano je prijevodno rješenje *jedni drugima*, koje u ovom kontekstu nije dovoljno precizno. Ostavlja se dojam da nastavnici kao jedinice moraju nešto priznavati drugim nastavnicima, primjerice u njihovu kolektivu, što je izraženo u odgovoru C. Riječ je o vrlo suptilnoj nezgrapnosti, a ipak dovoljnoj da onemogući jasno razumijevanje rečenice kada se traže odgovori na ovako precizna pitanja. Ovakvih se nepreciznosti osobito treba čuvati u neuronskim strojnim prijevodima kod kojih je već uočeno da se češće javljaju takve teškoće, premda se općenito smatra da su točniji i tečniji.

Zašto je toliko ispitanika na šesto pitanje odabralo odgovor D: *(nastavnici se boje)...u nastavi iskušati nešto novo?* Ako se vratimo na Primjer 5, možda se može pronaći djelomičan razlog. U njima autorica detaljno nabraja što bi sve nastavnici mogli iskušati – njezin je fokus zapravo na zajedničkom iskušavanju *pripreme nastavnih materijala, alternativnih metoda i razvoja oblika komunikacije*, a ne na tome da je to nešto ново – ali ispitanici su mogli pomisliti da je upravo to odgovor na ovo pitanje. Može se ponuditi objašnjenje da je naprosto riječ o jednom od težih pitanja, koja su ionako namjerno koncipirana tako da se njima pokušavaju zbuniti kandidat, kao i da je njihov zadatak upravo pokazati dubinsko razumijevanje teksta probijanjem kroz te zamke.

Drugo vjerojatno objašnjenje glasi da je mjesto na kome se točan odgovor krio (navedeno u Primjeru 6) prevedeno slabije razumljivo. Opis nastavnog sata kao „immer wieder ein Kampf um Anerkennung, ein Ringen um Aufmerksamkeit und Disziplin“ (*uvijek je on iznova borba za priznanjem, pažnjom i disciplinom*) na njemačkom jeziku jasno podrazumijeva da se nastavnici bore za priznanje *učenika*, te da traže pažnju i disciplinu *učenika* i to nije potrebno eksplicitno ponoviti. U strojnom prijevodu „lekcija“ je „uvijek iznova borba za priznanje, borba

za pažnju i disciplinu“, ali iz toga se ne iščitava dovoljno jasno da se nastavnici boje da ih učenici neće dobro prihvatiti.

U konvencionalnom je prijevodu ljudski prevoditelj zaobišao obje opisane problematike – *untereinander* uopće nije preveo, ali je stoga izrazom „kada bismo si mi nastavnici“ pokrio značenje onoga što se u izvorniku tim prilogom htjelo izraziti. Kod *borbe za priznanje, pažnju i disciplinu* ljudski je prevoditelj odabrao iste lekseme kao strojni, ali je dodao zamjenicu *on* kako bi stvorio bolju poveznicu između tog završnog dijela rečenice s prvim, kada se spominje *nastavni sat*. Iz toga je vidljivo da alat za strojno prevođenje nije nužno pogriješio u leksičkom odabiru riječi, no konvencionalno prevođenje uključuje i donošenje odluka o tome kako što koherentnije i jasnije oblikovati tekst koji na ciljnom jeziku ne može uvijek samo pratiti strukturu izvorne rečenice.

Zamijetili su to i ispitanici – u svojim komentarima čak su i oni ispitanici koji su pokazali generalno pozitivne stavove prema strojnom prevođenju ili se strojnim prevođenjem često služe, prijevodu zamjerali „doslovnost“, „prevođenje riječ po riječ“ ili su komentirali da prijevod „nije u duhu hrvatskog jezika“. Neki od tih komentara, koji pokazuju da su i pozitivno nastrojeni ispitanici ipak svjesni nedostataka u strojnim prijevodima, glase:

I10: *Smatram da su korisni pri prevođenju nepoznatih riječi u tekstu, ali ne i za prevođenje čitavih rečenica i tekstova jer se onda često gubi smisao rečenice.*

I13: *Strojni su prijevodi korisni na razini razumijevanja, ali nikako na razini upotrebljivosti, dakle iako nam je rečenica jasnija i dalje je moramo prilagoditi duhu hrvatskog jezika.*

I50: *Smatram kako su ovakvi alati korisni u prevođenju pojedinih riječi ili učestalih fraza, ali u prevođenju kompletnih rečenica i/ili tekstova ponekad mogu biti netočni, jer se najčešće radi o doslovnom prevođenju jedne po jedne, što nije uvijek pouzdano jer svaki jezik ima svoju logiku poretka riječi u rečenici, slaganja dijelova rečenice, prenesenog značenja pojedinih fraza i slično. (...)*

5.2.3 Proporcionalnost broja točnih odgovora na testu razumljivosti s ocjenama razumljivosti

Za postavljeno istraživačko pitanje o tome koliko su ocjene razumljivosti ispitanika proporcionalne broju točnih odgovora zabilježeno je da je šezdeset i devetero ispitanika (43,1 %) prijevod ocijenilo proporcionalno rezultatima koje su pokazali na testu razumijevanja. To znači da su izlazne podatke ocijenili najvišim ocjenama te ujedno imali više od četiri točna odgovora od sedam postavljenih pitanja ili da su prijevod ocijenili nižim ocjenama, ali da su isto tako imali samo do dva točna odgovora od mogućih sedam. Kod četrdeset i devetero ispitanika (30,6 %) slučaj je bio obrnut, a kod preostalih četrdeset i dvoje ispitanika (26,25 %) zabilježena su manja odstupanja u broju točnih odgovora i ocjene koju su dodijelili prijevodu. Od tih četrdeset i devetero ispitanika njih je desetero točno odgovorilo na sva pitanja u testu, i od toga je osmero pretpostavilo da su odabrali većinu točnih odgovora, što svjedoči o tome da su prilično jednostavno pronašli točne odgovore, unatoč tome što su razumljivost prijevoda ocijenili nižim ocjenama. K tome je jedan od tih ispitanika na sva pitanja točno odgovorio u samo dvije minute, dok je ostalima trebalo pet do jedanaest minuta.

Ovi podaci svjedoče o tome da rezultati subjektivne procjene ne koreliraju s rezultatima testa razumijevanja, ali i da su rezultati metode razumijevanja valjaniji jer se proučava stvarna uspješnost na testu. Subjektivni stavovi nisu jednako pouzdani: zabilježeno je, primjerice, četrnaestero ispitanika sa šest točnih odgovora na pitanjima u testu koji su prijevodu dodijelili niske ocjene razumljivosti, uglavnom ocjenu 2, a poneki 3. Od tih četrnaestero ispitanika, dvanaestero se čak izjasnilo i da su prilično sigurni u to da su odabrali točne odgovore na sva pitanja, a od toga je troje sve to tako uspješno napravilo u samo dvije minute. Ostali su za rješavanje uzeli tri do deset minuta, a samo je jedan od tih ispitanika na test potrošio trideset minuta. Sudeći po ocjenama ispitanika, reklo bi se da tekst nije osobito kvalitetno preveden, a sudeći po rezultatima testa, vidimo da su ispitanici ipak iščitali mnogo značenja.

Valja posebno proučiti i primjer ispitanika koji je na testu imao pet točnih odgovora, a prijevod je ocijenio „potpuno nerazumljivim“ (ocjena 1) te je smatrao da je odabrao većinu netočnih odgovora. I on je rješavanju testa posvetio samo dvije minute, a jedan je od one nekolicine ispitanika koji su iskazali općenito vrlo negativan stav prema strojnom prevođenju. Iz njegove se subjektivne procjene razumljivosti mora pretpostaviti da je tekst posve nerazumljiv, istraživač nema uvid u razloge zašto je ispitanik odabrao tako nisku ocjenu i mora je kao takvu prihvatiti, čak i ako uvidi da je riječ o ispitaniku s općenito negativnim stavovima

prema strojnom prevođenju. Metoda testa razumijevanja u potpunosti mijenja taj sud: ispitanik je čak i s tako malo uloženog truda točno odgovorio na veći dio pitanja. Njegova se procjena kvalitete sada doista može smatrati posljedicom loših iskustava sa strojnim prevođenjem, a stvarno razumijevanje iščitati iz broja točnih odgovora na testu.

Postoje, dakako, i suprotne situacije: ispitanici koji ni na jedno pitanje nisu točno odgovorili ili su odgovorili na samo jedno ili dva pitanja točno, a prijevodu su dali visoke ocjene 4 ili 5. Takva je disproporcija zabilježena kod dvadeset i jednog ispitanika. Razlozi zašto bi netko tko je prijevod procijenio razumljivim imao vrlo malo ili nijedan točan odgovor mogu biti mnogostruki. Primjerice, jedan od dva ispitanika bez ijednog točnog odgovora pritom je i sam smatrao da nije nimalo siguran u odabir svojih odgovora, a također je izdvojio najmanje vremena za rješavanje od svih ispitanika – tek minutu. Može se smatrati da se neki ispitanici ne potrudu dovoljno oko rješavanja testa razumijevanja, iako su spremni priznati da im tekst u načelu zvuči razumljivo. Također se ponovno u obzir mora uzeti ljudski faktor jer ljudi nekada jednostavno precijene svoje mogućnosti, a kao što je već navedeno, i rješavanje testova razumijevanja doista je već vještina za sebe, kojom ne vladaju svi ispitanici jednako.

Na temelju navedene analize odgovor na istraživačko pitanje *Jesu li ocjene pri procjeni razumljivosti teksta proporcionalne broju točnih odgovora na testu razumijevanja?* glasi da su ocjene razumljivosti tek djelomično proporcionalne rezultatima na testu razumljivosti. To znači da se pri vrednovanju istog teksta, od istih ispitanika i u istim uvjetima, ne postižu isti rezultati metodom subjektivne procjene i metodom testa razumijevanja. U ranijim istraživanjima s metodom testova razumijevanja može se uočiti slična tendencija. U istraživanju koje su proveli Lieve Macken i Iris Ghyselen (2018: 123) vidimo da su tri prijevoda Google Prevoditelja dobila razmjerno sličnu prosječnu ocjenu razumljivosti cjelokupnog teksta kao u ovom istraživanju ali rezultati testa razumijevanja iznenadili su istraživače. Svaki je od ukupno devedeset i devet ispitanika pročitao dva teksta koji su mogli biti ili konvencionalno ljudski ili strojni prijevod, što su prvo pokušali razlučiti, zatim su prijevodu dali ocjenu razumljivosti i na koncu odgovarali na set pitanja u testu razumijevanja. Strojni su prijevodi napravljeni alatima Google Prevoditelj te DeepL, i samo u prvom tekstu na testu je razumijevanja najbolje rezultate imao ljudski prijevod. Test razumijevanja za drugi tekst najbolje su riješili ispitanici koji su čitali prijevod napravljen alatom DeepL, a na trećem tekstu najuspješniji su bili oni koji su čitali prijevod Google Prevoditelja (2018: 124). Subjektivne ocjene nisu se poklapale s tim rezultatima: ljudski prijevod dobio je najviše prosječne ocjene za sva tri teksta: ocjenu 4,1 za prvi i drugi tekst te

4,0 za treći tekst. Google Prevoditelj dobio je ocjenu 3,5 za prvi i drugi tekst te ocjenu 3,1 za treći tekst (isto, 123), što je usporedivo s ocjenom 3,6 za prijevod iz ovog istraživanja. U oba istraživanja subjektivne ocjene ne mogu se direktno povezati s razinom uspješnosti na testu razumijevanja. U potpoglavlju 5.6 detaljno će se stoga usporediti te dvije metode i objektivnost rezultata koje se njima dobivaju, zajedno s metodom analize pogrešaka.

5.3 Metoda analize pogrešaka: rezultati i diskusija (IP4, IP5)

Sljedeća potpoglavlja posvećena su rezultatima provedene analize pogrešaka: prvo će se na primjerima pokazati kakve su vrste pogrešaka zabilježene u predmetnom strojnom prijevodu, a potom će se kvalitativno tumačenje i rasprava usmjeriti na IP4 (*Ovise li ocjene pri procjeni razumljivosti pojedinih segmenata teksta o broju jezičnih pogrešaka u njima?*) te na IP5 (*Utječu li pogreške u leksičkoj točnosti bitnije na nerazumijevanje prijevoda od ostalih pogrešaka?*) i pokušat će se dati odgovori na postavljena istraživačka pitanja.

5.3.1 Kategorije anotiranih pogrešaka (rezultati i analiza primjera)

Prvo valja reći da se rezultati ove analize pogrešaka ne mogu promatrati kao apsolutna istina jer je analiza pogrešaka, kao što su i recentnija istraživanja zamijetila (v. Klubička, Toral i Sánchez-Cartagena 2018: 206), sve više prihvaćena kao stvar mišljenja odabranog anotatora. Također treba uzeti i obzir da ni u teoriji nije moguće jasno definirati što je uopće pogreška i koje sve rečenične elemente neka pogreška zahvaća, što u praksi znači da bi neki drugi anotator pogreški zacijelio drukčije analizirao predmetni tekst. Na rezultate ove metode stoga naj snažnije utječe odabir anotatora, koji je u ovom slučaju bio profesionalni prevoditelj i lingvist s iskustvom u redakturi strojnih prijevoda. Anotator se u analizi pogrešaka odlučio voditi mišlju da označava samo one pogreške koje bi apsolutno zahtijevale intervencije u prijevodu kada bi se on redigirao za objavu, premda bi u stvarno slučaju naknadnog teksta uvrstio i više izmjena.

Od ukupno trideset i osam segmenata strojnog prijevoda u jedanaest segmenata nije zabilježena nijedna pogreška. Anotator je zabilježio šest kratkih segmenata, četiri srednje duga segmenta i jedan dugi segment u kojemu nije vidio pogreške, od čega je ovdje izdvojen jedan kratki, Primjer 7, i jedan duži segment, Primjer 8:

Primjer 7. *Volim svoj posao.*

Referentni prijevod: *Volim svoj posao.*

Izvornik: *Ich mag meinen Beruf.*

Primjer 8. *Ova prva rečenica nekako nije samo kratka i jezgrovita informacija o mom poslu, već priznanje.*

Referentni prijevod: *Nekako mi ta prva rečenica ne zvuči samo kao kratka i jasna informacija o mom zanimanju, već kao priznanje.*

Izvornik: *Irgendwie ist dieser erste Satz nicht nur die kurze und knappe Information über meinen Beruf, sondern ein Geständnis.*

Čak i kod ovih kratkih i razmjerno jednostavnih segmenata kao u Primjeru 7 može se raspravljati o načinima definiranja pogreške i o tome koliko analiza dubinski ovisi o anotatoru. Njemačka se imenica *Beruf* (koja se javlja i u Primjeru 7 i u Primjeru 8) u Njemačko-hrvatskom univerzalnom rječniku definira kao: „zanimanje, zvanje, poziv, profesija; posao“ (Hansen-Kokoruš i dr. 2005: 223). Anotator je odabirom ekvivalenta *posao* bio zadovoljan u oba primjera. S druge strane, u konvencionalnom prijevodu u drugom je slučaju odabran ekvivalent *zanimanje*, a moguće je zamisliti i da bi odgovarao neki od drugih ponuđenih ekvivalenata. Iako se odabrani anotator vodio maksimom najmanjeg broja neophodnih ispravaka, navedeni primjeri služe kao prikaz toga koliko samo definiranje prijevodne pogreške ovisi o jezičnom osjećaju sudionika istraživanja, od anotatora pogrešaka, ali i prevoditelja koji izrađuju referentne prijevode.

Tamo gdje jest uočio pogreške, anotator ih je razdijelio u šest kategorija: izostavljene riječi, pogrešan leksički odabir, pogrešan morfosintaktički oblik riječi, pogreška u ortografiji, pogreška u redosljedu riječi, te otvorena kategorija „ostalo“ kojom su obuhvaćene ostale vrste pogrešaka. Kategorije će se u nastavku oprimirati, svaki put s fokusom samo na jednoj kategoriji (naznačenih masno otisnutim dijelovima), premda se u mnogima od primjera navedenima u nastavku može uočiti i više od jedne pogreške.

A. Kategorija pogrešaka – **izostavljene riječi**:

Primjer 9. *Naravno, mogli bismo uvijek iznova ukazivati na ostale, žaliti se na ono što nam se radi i pozivati se na zadatke kojima teško možemo upravljati - jer često moramo preuzeti i **o zadatke roditeljstva**.*

Referentni prijevod: *Što mi nastavnici možemo učiniti da okrenemo stvari nabolje? Naravno da bismo mogli upirati prstom u druge, žaliti se na sve ono što nam je učinjeno i ukazati na to koliko je teško savladati sve naše zadatke – jer često, naime, moramo preuzeti i **odgojnu ulogu roditelja**.*

Izvornik: *Natürlich könnten wir immer wieder auf die anderen zeigen, darüber klagen, was uns angetan wird und auf unsere kaum zu bewältigenden Aufgaben verweisen – denn wir müssen nämlich oft auch **Erziehungsaufgaben** der Eltern übernehmen.*

Primjer 9 služi kao primjer za izostavljene riječi, kojih je anotator zabilježio ukupno šest. Izvorni izraz *Erziehungsaufgaben der Eltern* doslovce bi se mogao prevesti kao „odgojni zadaci roditelja“, a u referentnom je prijevodu to idiomatičnije izraženo kao *odgojna uloga roditelja*. Strojni prijevod na tom je mjestu neprecizan i izostavljena je informacija o vrsti „zadataka roditeljstva“ kakve nastavnici često moraju preuzeti. Ovaj primjer prikazuje i da anotator kategoriju „izostavljenih riječi“ nije doslovno shvatio kao nešto što se mora odnositi na jedinicu riječi, već da se granice pogrešaka mogu katkad odnositi i na manje ili veće jedinice, u ovom slučaju na jedan član njemačke složenice. To je legitiman postupak anotatora da kategorije definira onako kako smatra da je to u konkretnom slučaju najispravnije, a važno je samo da u tome bude dosljedan. Isto tako je neki drugi anotator navedeni primjer mogao kategorizirati, primjerice, kao leksičku pogrešku jer se pogreška zapravo javlja na razini jednog njemačkog leksema (*Erziehungsaufgaben*), što pokazuje koliko je analiza pogrešaka dubinski ovisna o jezičnom osjećaju svakog pojedinog anotatora. Ta subjektivnost pri anotiranju jedna je od bitnijih nedostataka metode analize pogrešaka.

B. Kategorija pogrešaka – leksičke pogreške:

Zabilježene su i dvadeset i tri leksičke pogreške. Premda će govora o utjecaju leksičkih pogrešaka na (ne)razumijevanje teksta biti u potpoglavlju 5.3.3, Primjer 10 pokazuje o kakvim je tu pogreškama riječ:

Primjer 10. *Postoje razne **studije iz Pise**, koje više puta dokazuju da su naši studenti u međunarodnoj usporedbi tek osrednji.*

Referentni prijevod: *Iskaču neke **PISA-ine studije** koje neprestano dokazuju da su naši učenici u međunarodnoj usporedbi tek osrednji.*

Izvornik: *Da sind die diversen **Pisa-Studien**, die immer wieder belegen, dass deutsche Schüler im internationalen Vergleich gerade mal nur Mittelmaß sind.*

Strojni prevoditelj nije prepoznao da se izraz *Pisa-Studien* ne odnosi na talijanski grad Pisu u kojoj bi tobože bile provedene razne studije, već da je riječ o međunarodnim testiranjima i istraživanjima procjena znanja i vještina učenika poznatih pod kraticom PISA (*Programme for International Student Assessment*). Kod utjecaja leksičkih pogrešaka na razumijevanje treba uzeti u obzir dvije stvari: ova bi se pogreška svakako mogla označiti kao gruba pogreška, ali s druge bi je strane čitatelj koji je upoznat s tematikom lako mogao prozreti i ipak povezati što na tom mjestu doista piše u izvorniku. Tako postoji mogućnost da u nekim slučajevima čitatelj i unatoč pogreški ispravno razumije ovaj segment. Drugo je važno obilježje ove vrste pogreške to da bi se vrlo lako mogla ispraviti kada bi se radila redaktura ovog teksta, lakše nego, primjerice, već opisane izostavljene pogreške koje se mogu uočiti samo pomnom usporedbom prijevoda i izvornika.

C. Kategorija pogrešaka – morfosintaktičke pogreške:

U izlaznim je podacima zabilježeno trinaest morfosintaktičkih pogrešaka, koje se učestalo nastavljaju na neke druge pogrešne odabire, kao u Primjeru 11:

Primjer 11. *Ne, kad sam studirao "za učitelja", bio sam pun ideala.*

Referentni prijevod: *Ne, kada sam upisala studij da postanem nastavnica, bila sam puna ideala.*

Izvornik: *Nein, als ich damals "auf Lehramt" studierte, steckte ich voller Ideale.*

Na prvi pogled, promatrajući samo izlazne podatke, ovaj segment ne sadrži morfolosintaktičke pogreške. Ni na drugi pogled, u usporedbi izlaznih podataka s izvornikom, ovaj segment ne sadrži morfosintaktičke pogreške. No anotator pred sobom ima cjelovit izvornik, i zna da drugi segment glasi: *Ich bin Lehrerin und unterrichte Deutsch und Geschichte.* odnosno u referentnom prijevodu: *Nastavnica sam i predajem njemački te povijest.* Autorica je teksta žena. To se ne vidi u gramatičkim oblicima drugih njemačkih riječi (*studierte, steckte voller Ideale*) osim imenica s nastavcima koji sugeriraju ženski rod (poput *Lehrerin*), ali u hrvatskim oblicima mora biti jasno naznačeno. S time se složio i anotator koji je svaki oblik riječi u muškom obliku smatrao pogrešnim morfološkim odabirom (*studirao, bio, pun*). Ispitanici, koji pred sobom nisu imali ni izvornik ni referentni prijevod, do kraja istraživanja nisu mogli biti posve sigurni u rodnu odrednicu autora teksta jer je strojni prijevod po tom pitanju neusklađen. Svi su glagolski oblici u tekstu prevedeni u muškom obliku, premda drugi segment i u strojnom prijevodu glasi: *Ja sam učiteljica i predajem njemački jezik i povijest.*

I ovaj segment pokazuje kako kategorije pogrešaka daleko mogu nadmašivati samo kategoriju riječi te da je uvijek u obzir potrebno uzeti tekst kao cjelinu. Na ovom je mjestu važno zabilježiti da u ovom (neuronskom) strojnom prijevodu nisu zabilježene najklasičnije i „očitiije“ morfosintaktičke pogreške kakve su se znale javljati u starijim strojnim prijevodima te kakve se možda očekuju unutar kategorije morfosintaktičkih pogrešaka, ili kakve se čak znaju greškom pronaći u ljudskim prijevodima. To potvrđuje navode iz dosadašnje literature da prijevodi neuronskih sustava zvuče sve „tečnije“; istraživači bilježe mnogo manje morfosintaktičkih pogrešaka nego na statističkim sustavima, a pogreške su zastupljenije u drugim kategorijama, poput izostavljenih riječi. Od zabilježenih trinaest morfosintaktičkih pogrešaka šest se odnosilo na oblike glagola kao u Primjeru 11. S druge strane, u Primjeru 12 prikazana je morfosintaktička pogreška koja je nastala pogrešnim odnosno neidiomatskim prijevodom njemačke glagolske rekcije:

Primjer 12. *Nekim studentima to nije dovoljno i pitaju o značenju cijelog postupka ispitivanja.*

Referentni prijevod: *Neki učenici smatraju da to nije dovoljno i pitaju koji je uopće smisao čitavog upitnika.*

Izvornik: *Manchen Schülern ist das zu wenig und sie fragen nach dem Sinn der ganzen Frageaktion.*

Anotator je zabilježio neispravnu uporabu prijedloga *o* u frazi *pitaju o značenju*. Pritom ne računa oblik riječi *značenju* kao pogrešan, premda bi se i on trebao promijeniti ako bi se pri redakturi teksta prijedlog mijenjao te tada zahtijevao drugi padež, već kao morfosintaktičku pogrešku broji samo prijedlog. Unatoč tome što se radi o kratkoj riječci, iz referentnog je prijevoda razvidno da je ne bi nužno bilo tako lako ispraviti u postupku redakture prijevoda, nego da bi se potencijalno morala preurediti čitava rečenica da bi na ciljnom jeziku zvučala prirodno.

D. Kategorija pogrešaka – **ortografske pogreške:**

Segment u Primjeru 13 sadržava najviše pogrešaka, a među njima je anotator zabilježio i jednu ortografsku. U hrvatskom se jeziku, kao i u njemačkom, znak navoda otvara dolje, no strojni prevoditelj pogriješio je na mjestu postavljanja navodnika. Ta se pogreška stalno iznova zamjećuje u prijevodima Google Prevoditelja, što se pripisuje tome što Google i dalje prevodi s engleskim kao jezikom posrednikom, pa u tom međukoraku preuzima englesku ortografiju. U tekstu se dvaput javljaju rečenice s navodnim znakovima, i oba je puta došlo do iste pogreške, i to su jedine dvije ortografske pogreške anotirane u tekstu.

Primjer 13. *Rečenica "Ja sam učitelj" uvijek mi se čini kao priznanje, kao priznanje nečemu negativnom, otprilike slično priznanju poroku poput pušenja ili još gore.*

Referentni prijevod: *Rečenica „Ja sam nastavnica“ uvijek mi zvuči kao neko priznanje, kao da priznajem nešto negativno, otprilike kao da imam neki porok, recimo da pušim ili što još gore.*

Izvornik: *Der Satz „Ich bin Lehrerin“ kommt mir stets wie ein Bekenntnis vor, wie ein Bekenntnis zu etwas Negativem, ungefähr ähnlich dem Bekenntnis zu einem Laster wie dem Rauchen oder Schlimmerem.*

E. Kategorija pogrešaka – **pogrešan redoslijed riječi:**

Predzadnju kategoriju pogrešaka čini redoslijed riječi u rečenici. Neuronski su sustavi za strojno prevođenje poznati po tome da drastično manje griješe u redoslijedu riječi, pa je tako i u ovom prijevodu došlo do samo jedne takve pogreške, prikazane u Primjeru 14:

Primjer 14. *Na kraju, uvijek su učitelji ti koji ne mogu stvari kontrolirati, koji se ne trude dovoljno ili su čak neprikladni za svoju profesiju.*

Referentni prijevod: *Na kraju su uvijek nastavnici ti koji navodno nemaju stvari pod kontrolom, koji se premalo trude ili čak uopće nisu sposobni za obavljanje njihova zanimanja.*

Izvornik: *Letztendlich sind es immer die Lehrer, die die Sache nicht in den Griff kriegen, sich zu wenig Mühe geben oder gar für ihren Beruf nicht geeignet sind.*

Primjer 14 upućuje na to kako se isti elementi mogu ubrajati u više kategorija pogrešaka. Prijevodno rješenje za njemački frazem *die Sache nicht in den Griff kriegen* („nemati stvari pod kontrolom“, referentni prijevod), koji u strojnom prijevodu glasi „ne mogu kontrolirati stvari“, nije optimalan i različiti anotatori mogli bi ga svrstati u različite kategorije. No osobito je uočljivo to što je u strojnom prijevodu preuzeta njemačka sintaksa zavisnih surečenica – glagol je postavljen na zadnje sintaktičko mjesto. Premda hrvatski jezik ima slobodniju sintaksu od njemačke, u ovom bi slučaju objekt trebao doći iza glagola, tako da je ovo jasan primjer pogreške u redoslijedu riječi.

F. Kategorija pogrešaka – **ostalo:**

Na kraju klasifikacije imamo najotvoreniju kategoriju svih *ostalih* pogrešaka. One se ne mogu klasificirati u neke od dosad navedenih. Kategorija služi tomu da omogući anotatoru da označi mjesta koja smatra pogrešnima, ali za koja nije sasvim načisto kako bi ih trebao kategorizirati,

kao u Primjeru 15. Stoga anotator uz svaku takvu pogrešku treba priložiti pojašnjenje. U ovom je strojnom prijevodu anotator zabilježio pet takvih pogrešaka.

Primjer 15. *Smiješno je da mi učitelji, koje **neprestano i neprestano ocjenjujemo i ocjenjujemo** druge i njihov rad, nismo u stanju pouzdano se nositi s kritikama ili prosudbama o sebi.*

Referentni prijevod: *Ipak je smiješno da mi nastavnici, koji **neprestano (pr)ocjenjujemo** druge i njihov rad, nismo u stanju suvereno se nositi s kritikom ili mišljenjima o nama samima.*

Izvornik: *Es ist schon komisch, dass wir Lehrer, die wir **stets und ständig** andere und ihre Arbeit **bewerten und beurteilen**, nicht in der Lage sind, souverän mit Kritik oder Urteilen über uns selber umzugehen.*

Njemačka kolokacija *stets und ständig* okamenjena je fraza koja se sastoji od dviju bliskoznačnica: *stets* čije je denotativno značenje „uvijek, stalno, neprestano, vazda“ (prema Hansen-Kokoruš i dr. 2005: 1628) i riječi *ständig* denotativnog značenja „stalan, trajan, neprekidan, neprestan, vrlo čest“ (isto, 1613). U hrvatskom je jeziku frazu potrebno zamijeniti istovrijednom višечlanom frazom ili jednom riječju koja bi obuhvatila njezino značenje, kao što je učinjeno u referentnom prijevodu s riječju *neprestano*. Sustav za strojno prevođenje riječi je preveo zasebno, a pogreška je nastala tako što su za obje riječi odabrana ista rješenja u izlaznim podacima (*neprestano i neprestano*), što bi se moglo smatrati pragmatičkom ili prijevodnom pogreškom.

Ista se pogreška ponavlja i u nastavku, gdje se pojavljuju dva glagola: *bewerten und beurteilen*. U ovom slučaju glagoli ne čine okamenjenu frazu, već autorica želi naglasiti svoju poantu ponavljanjem dviju bliskoznačnica: glagola *bewerten* denotativnog značenja „ocijeniti/ocjenjivati, suditi, procijeniti/procjenjivati, vrednovati“ (isto, 246) i glagol *beurteilen* denotativnog značenja „ocijeniti/ocjenjivati, prosuditi/prosuđivati, procijeniti/procjenjivati (isto, 243). U nekim bi se kontekstima za ove glagole mogli odabrati isti prijevodni ekvivalenti, ali ne i kada se pojavljuju jedan za drugim, kao u izlaznim podacima: „ocjenjujemo i ocjenjujemo“. U referentnom prijevodu vidimo mogućnosti prevoditeljske kreativnosti jer je ljudski prevoditelj sporni dio preveo kao: *neprestano (pr)ocjenjujemo (...)*.

U oba ova slučaja anotator je računao jednu pogrešku za kategoriju ostalo, uz obrazloženje „da ih ne smatra klasičnim leksičkim pogreškama jer su za same riječi, gledano izvan konteksta rečenice, odabrani dobri leksički ekvivalenti, ali oni pragmatički ne funkcioniraju u prijevodu“. Iako iz teorije prevođenja znamo da se zasebne riječi ne mogu *prevoditi* izvan konteksta, anotatorova argumentacija slijedi jasnu logiku koju je odredio sebi kao mjerilo, i informativnije je zadržati kategoriju ostalih pogrešaka za ovakve slučaje, nego prisiliti anotatora da se odluči za kategoriju koja mu se ne čini u potpunosti ispravnom.

5.3.2 Utjecaj broja jezičnih pogrešaka na ocjene razumljivosti (IP4, rezultati i diskusija)

Veliko je pitanje povezano s metodom analize pogrešaka znači li veći broj pogrešaka ujedno i da je tekst čitateljima nerazumljiviji? Analizom segmenata zamjećuje se da se u većini rečenica u tekstnom materijalu ne javlja osobito velik broj pogrešaka. Premda su neke od njih grublje od drugih, ipak ih nema toliko mnogo. Od ukupno trideset i osam segmenata:

- u jedanaest segmenata nije pronađena nijedna pogreška;
- u dvanaest je segmenata uočena samo jedna pogreška;
- u jedanaest segmenata uočene su dvije pogreške.

To znači da su samo u dva segmenta anotirane tri pogreške, u jednom četiri pogreške i u jednom šest pogrešaka. Čak i kada se u obzir uzme duljina segmenata, u dvadeset je segmenata pronađeno manje od 20 % pogrešno prevedenih elemenata u odnosu na broj riječi. Po istom je kriteriju uočeno šest segmenata u kojima ima od 20 % do 40 % pogrešnih elemenata, i samo jedan koji se smatra stopostotno pogrešno prevedenim.

Suprotno očekivanjima, od segmenta koji bi bio stopostotno pogrešno preveden, u Primjeru 16 vidimo da on nije nerazumljiv. Zapravo se strojni prijevod sastoji od jedne riječi, a anotator je zabilježio jednu pogrešku. Čitajući ovaj strojni prijevod bez usporedbe s referentnim prijevodom ili izvornikom, reklo bi se da je prijevod neupitno ispravan, ali anotator je, usporedivši prijevod s izvornikom (koji se sastoji od dvije riječi), zaključio da je došlo do pogreške izostavljene riječi.

Primjer 16. *Zašto ø?*

Referentni prijevod: *A zašto?*

Izvornik: *Wieso eigentlich?*

Ljudski je prevoditelj odabrao ovaj segment prevesti kao: „A zašto?“ Vidimo da su i prevoditelj i anotator smatrali da uz glavnu informaciju treba prenijeti dodatnu, pragmatičku nijansu značenja koja, premda je doista suptilna, zaista postoji u njemačkom izvorniku koji prenosi određeno čuđenje. U njemačkom je jeziku riječ *eigentlich* definirana kao modalna čestica sa značenjem „fokusiranja bitnog ili uvođenja nove teme“ (Kresić Vukosav i Batinić Angster 2019: 12), a u upitnim se rečenicama, s upitnom riječju, često prevodi kao *zapravo* ili *nego* (Kresić i Batinić 2014: 113). U referentnom prijevodu fokusiranje na bitno postiglo se s riječju *a* na početku pitanja. Bez tog dodatka ona nije posve vjerno prevedena. S druge strane ispitanici kojima je predstavljen samo strojni prijevod, bez izvornika ili referentnog prijevoda, ne mogu iz segmenta u Primjeru 16 vidjeti pogrešku, ona je neprozirna jer segment funkcionira posve samostalno i djeluje razumljivo. Gotovo čudi što ukupna ocjena razumljivosti dodijeljena ovom segmentu nije i viša od 4,1, što je zapravo prosječna ocjena razumljivosti.

Čak ni analizom segmenata kod kojih su ispitanici lakše mogli uočiti jezične pogreške nije uočena jasan suodnos između broja pogrešaka i ocjene razumljivosti segmenta. Ni u segmentima s većim brojem pogrešno prevedenih elemenata nije zamijećeno da je to utjecalo na negativniju percepciju razumljivosti: tri su takva segmenta svejedno ocijenjena u rangju prosječne ocjene, a dva su čak bolje ocijenjena od prosjeka (4,8 i 4,9, odnosno vrlo blizu najvišoj mogućoj ocjeni). Da broj pogrešaka nije nužno povezan s razumljivošću, vidi se iz Primjera 17:

Primjer 17. *Drugo, studenti su sigurno naučili kako kritizirati.*

Referentni prijevod: *Učenici zatim moraju naučiti kako izreći kritiku.*

Izvornik: *Die Schüler müssen zweitens gelernt haben, wie man Kritik übt.*

Izbor rješenja *studenti* za njemački *Schüler* (*učenici*, prema referentnom prijevodu) nije točan i ne odgovara kontekstu srednjoškolskog obrazovanja. Ipak, sudeći po ocjeni segmenta, ta pogreška nije ispitanicima pretjerano omela razumijevanje, što se može objasniti time da su već iz konteksta mogli shvatiti da se tekst odnosi na srednje škole (v. i Primjer 19 u potpoglavlju 5.3.3 za daljnje objašnjenje).

Drugu pogrešku čini nespretan prijevod sintagme da su *studenti* (ili – *učenici*) *sigurno naučili kako kritizirati*. Njemački modalni glagol *müssen* ima denotativno značenje *morati*, pa se u referentnom prijevodu ljudski prevoditelj odlučio za rješenje: *učenici moraju naučiti kako izreći*

kritiku (da bi funkcioniralo vrednovanje nastavnika koje su uvele neke škole). Značenje modalnog glagola u drugim bi se kontekstima moglo raspisati riječju *sigurno*, ali ovdje to ne bi bilo primjereno rješenje. Unatoč tome, Primjer 17 ispitanici su ocijenili prosječnom ocjenom 4,3, što je više od prosjeka.

Zapravo je zabilježen samo jedan segment (Primjer 18) koji i sadrži veći broj pogrešaka i dobio je lošu prosječnu ocjenu razumljivosti, 2,3, što je poprilično ispod prosjeka:

Primjer 18. *Pripremaju li učitelji predmetne nastave zajedno?*

Referentni prijevod: *Pripremaju li predmetni nastavnici zajednički lekcije?*

Izvornik: *Bereiten Fachlehrer gemeinsam Unterrichtseinheiten vor?*

Njemačka imenica *Fachlehrer* označava *predmetne nastavnike* (prema Hansen-Kokoruš i dr. 2005: 1044), a ne *učitelje*, a *Unterrichtseinheit(en)* značenja je „nastavna jedinica“ ili „nastavna cjelina“ (isto, 1808). Ljudski prevoditelj odlučio se za također primjeren ekvivalent *lekcija*. U strojnom prijevodu ispitanicima je poteškoće u razumljivosti stvarao prijevod *predmetne nastave* koji, osim što nije točan, u hrvatskom jeziku također zvuči neprirodno i nejasno jer se o nastavi u ovom kontekstu prvenstveno govori u jednini. I u Primjer 18 riječ je o kratkom segmentu, za koje se načelno očekuje da će biti lakši, ali kada su ključni koncepti pogrešno prevedeni, ispitanicima je vrlo teško iz zadanog materijala složiti smislenu cjelinu. To bi moglo objasniti i lošu ocjenu u ovom konkretnom slučaju.

Zaključno, treba reći da u prevedenim segmentima nije zabilježeno osobito mnogo pogrešaka, ali također nije uočen ni jasan suodnos između broja pogrešaka i ocjene percipirane razumljivosti. Kao što je već navedeno, u segmentima se javljaju pretežno jedna do dvije pogreške, ako uopće. U tekstu od ukupno četrismo devedeset i osam riječi zabilježeno je pedeset pogrešaka: šest izostavljenih riječi, dvadeset i tri pogrešna leksička odabira, jedna pogreška u redosljedu riječi, trinaest pogrešnih morfosintaktičkih oblika riječi, dvije ortografske pogreške i pet ostalih pogrešaka (v. Graf 8 u idućem potpoglavlju za distribucija pogrešaka).

IP4 (*Ovise li ocjene pri procjeni razumljivosti pojedinih segmenata teksta o broju jezičnih pogrešaka u njima?*) odnosi se prvenstveno na broj pogrešaka, a kada se u obzir uzme broj pogrešaka u segmentima i to usporedi s ocjenama razumljivosti koje su ispitanici dodijelili tim segmentima, vidljivo je da je četrnaest segmenata (oko 37 %) ocijenjeno u skladu s brojem pogrešaka, odnosno obrnuto proporcionalno (gdje je više pogrešaka, ocjena razumljivosti je

niža; gdje je manje pogrešaka, ocjena je razumljivosti viša), a dvadeset i četiri segmenta (oko 63 %) ocijenjena su u neskladu s brojem pogrešaka.

Rezultati govore da odgovor na IP4 glasi da ne postoji jasan utjecaj broja pogrešaka i percipirane razumljivosti. Ovo je istraživačko pitanje u povijesti proučavanja vrednovanja strojnog prevođenja bilo relevantno, čak su i ranija istraživanja u hrvatskom kontekstu pokazala da je svladavanje morfoloških i sintaktičkih posebitosti jezika predstavljalo priličan izazov statističkim modelima strojnog prevođenja. Ipak, na temelju ovih rezultata može se zaključiti da to pitanje s uvođenjem neuronskog strojnog prevođenja gubi na značaju jer se u izlaznim podacima na hrvatskom jeziku bilježi sve manje morfosintaktičkih pogrešaka. Stoga je sljedećim istraživačkim pitanjem za razliku od broja pozornost usmjerena na utjecaj pojedinih vrsta pogrešaka na percipiranu razumljivost.

5.3.3 Utjecaj vrste jezičnih pogrešaka na ocjene razumljivosti (IP5, rezultati i diskusija)

Načelno se smatra da leksičke pogreške više utječu na nerazumljivost prijevoda. Katkad je dovoljno da samo jedan element u rečenici nije primjereno preveden jer on može biti glavni nositelj značenja i bez odgovarajućeg rješenja za njega ostatak rečenice nije smislen (kao što je bio slučaj i u Primjeru 18). Ipak, i to se shvaćanje temelji na starim, statističkim sustavima, a danas se situacija mijenja. Rezultati ovog istraživanja pokazuju da su leksičke pogreške, unatoč tome što ih i u ovom prijevodu ima najviše, rijetko kada posve neshvatljivo prevedene ili da se na mjestu njih javljaju potpuno nejasna rješenja, kao što je prije to nerijetko bio slučaj. Primjer 19 pokazuje kako se leksičke pogreške mogu i lako zanemariti, često kada treba samo shvatiti osnovnu misao teksta:

Primjer 19. *Da bi to uspjelo, moraju se dogoditi tri stvari: Prvo, **studenti** koji ocjenjuju moraju ostati anonimni.*

Referentni prijevod: *Da bi takvo vrednovanje funkcioniralo, moraju se dogoditi tri stvari: **Učenici** koji vrednuju nastavnike prvenstveno moraju ostati anonimni.*

Izvornik: *Damit sie funktioniert, muss dreierlei passieren: Erstens müssen die bewertenden **Schüler** anonym bleiben.*

U Primjeru 19 ponavlja se pogrešno leksičko rješenje za njemačku riječ *Schüler*, koja je prevedena kao *studenti* umjesto prikladnog *učenici*, kao što je bio slučaj i u Primjeru 17. No sudeći po ocjenama koje su ispitanici dodijelili ovom segmentu, sasvim je očito da ih ta pogreška nije zbunila – riječ je o najbolje ocijenjenom segmentu u čitavom prijevodu, svih ga je osamdesetero ispitanika koji su vrednovali raščlanjene segmente ocijenilo najvišom ocjenom 5. Taj se rezultat može različito protumačiti: može se smatrati da je značenjski tu ipak riječ o dvama ekvivalentima iz dovoljno bliskog semantičkog polja; ili da su ispitanici već iz konteksta bili svjesni da se u tekstu misli na srednjoškolske učenike; ili su ispitanici pak značenje čitavog segmenta, a možda i čitavog teksta, preslikali na fakultetske studente, ali im se rečenica svejedno i tako činila jasnou.

U jednu bi se ruku moglo smatrati da pogreške poput ove nisu grube leksičke pogreške, pa stoga ne utječu toliko izravno na nerazumijevanje, a u drugu da utjecaj leksičkih pogrešaka općenito nije veći od utjecaja bilo koje druge pogreške, odnosno da bi odgovor na IP5 glasio da leksička točnost kao kategorija pogrešaka nije bitnija od drugih kategorija. Tom tumačenju u prilog govori i Primjer 20, koji je također ocijenjen vrlo visokom prosječnom ocjenom 4,86:

Primjer 20. *I treće, samo dotični učitelj doživljava rezultate s kojima se onda može nositi kako želi.*

Referentni prijevod: *I, treće, samo dotični nastavnik smije saznati rezultate vrednovanja te s njima potom smije raditi što želi.*

Izvornik: *Und drittens erfährt allein die betroffene Lehrkraft die Ergebnisse, mit denen sie dann umgehen kann, wie sie will.*

Anotator je u ovom segmentu označio dvije leksičke pogreške. Već se i u pokazanim primjerima ponavljao slučaj kada je njemačka riječ *Lehrkraft* prevedena kao *učitelj*, što se u hrvatskom odnosi na „nastavnika u osnovnoj školi“ (prema Šonje 2000: 1291), a u srednjoškolskom bi kontekstu odgovarao ekvivalent *nastavnik* (Hansen-Kokoruš i dr. 2005: 1044). Međutim, kao i u primjeru *student/učenik*, na temelju odličnih se ocjena može pretpostaviti da ta pogreška nije osobito utjecala na percipiranu razumljivost segmenta, pogotovo što se te dvije pogreške dosljedno ponavljaju kroz čitav tekst.

Druga se leksička pogreška odnosi na njemački glagol *erfahren* koji je preveden kao *doživljava*. To jest jedno od njegovih značenja („dozna(va)ti, sazna(va)ti, spozna(va)ti;

doživjeti, pretrpjeti, iskusiti“; isto, 469), ali nije prikladno rješenje u ovom kontekstu. Ovom bi kontekstu najbolje odgovarao glagol *saznati*, kao u referentnom prijevodu, ili slična varijacija. Ipak, iz ocjena se ne može zaključiti da su ispitanici taj nedostatak u prijevodu ovog segmenta zamjerali sustavu.

U tekstu se javljaju i specifičnije leksičke pogreške, kao u Primjeru 21, ali i one se barem donekle nadovezuju na tekst, pa ponovno ta rješenja ne utječu drastično na slabo razumijevanje. Iz rezultata ovog istraživanja svakako se može zaključiti da se s godinama promijenila narav leksičkih pogrešaka, odnosno da se u prijevodima više ne javljaju posve nemotivirana rješenja.

Primjer 21. *Želio bih vam odgovoriti: budite strpljivi, barem je početak započet.*

Referentni prijevod: *Njima bih htjela odgovoriti: Strpite se, barem se nešto pokrenulo.*

Izvornik: *Ihnen möchte ich antworten: Habt Geduld, immerhin ist ein Anfang gemacht.*

I u Primjeru 21 anotirane su dvije leksičke pogreške. Prva se odnosi na prijevod riječi *ihnen*, gdje se može raditi o dativu njemačke zamjenice *sie* („oni“), kao što je slučaj u ovom segmentu, ili o obliku u jednini i iz poštovanja („vama“), kao što se pojavilo u izlaznim podacima. Iz konteksta rečenice jasno je da autorica kolumne pruža odgovor *učenicima* koji pitaju za značenje postupka vrednovanja nastavnika (*njima bih htjela odgovoriti*), što znači da je odabrano pogrešno leksičko rješenje.

Kod izraza *ein Anfang ist gemacht*, koji znači da *se nešto pokrenulo* ili da je nešto *započelo*, u strojnom se prijevodu našla neidiomska i pleonastična fraza *početak je započet*. Vjerojatno zbog prozirnosti značenja, ni to nije previše utjecalo na razumijevanje teksta, pa je prosječna ocjena ovog segmenta 3,7.

Ukupno su rezultati ovog istraživanja pokazali da su izlazni podaci prilično kvalitetni, pogotovo što su na temelju dosadašnjih istraživanja sa strojnim prevođenjem na hrvatski (s jezika koji nije engleski) očekivanja bila relativno niska. Ukupno je samo 10 % teksta klasificirano kao neka jezična pogreška, i nema rečenica koje vrve od potpuno nejasnih pogrešaka (v. poglavlje 2.1.2 za usporedbu s rezultatima starijih istraživanja). Od svih tih pogrešaka 4,8 % klasificirano je kao leksičke pogreške, no analizom materijala postaje jasno da većina njih potpada pod sitnije pogreške.

U ovom se istraživanju od anotatora nije tražilo da doda koeficijente pogrešaka ili na druge načine određuje njihovu težinu jer je shvaćanje težine vrlo individualno, ali naknadnom se analizom za neke pogreške može ustvrditi da nisu tako grube upravo zato što nisu osobito utjecale na ocjene razumljivosti. Također se učestalo ponavljaju iste pogreške (već navedeni *učitelj* i *studenti* umjesto *nastavnik* i *učenici*). Ozbiljnije se pogreške češće odnose na glagole (v. Primjere 17, 20, 21), ali ni tada ne moraju ostaviti toliko snažan otisak na ocjenu razumljivosti, kao u Primjeru 22:

Primjer 22. *Rad s djecom i mladima, prenošenje znanja i vrijednosti - nisam mogao smisliti bolji posao.*

Referentni prijevod: Rad s djecom i mladima, prenošenje znanja i vrijednosti – nisam mogla **zamisliti** bolje zvanje.

Izvornik: *Die Arbeit mit Kindern und Jugendlichen, die Vermittlung von Wissen und Werten – ich konnte mir keinen besseren Beruf vorstellen.*

Njemački glagol *vorstellen* prikladnije je u hrvatskom zamijeniti glagolom *zamisliti*, kao što je to učinjeno u referentnom prijevodu, dok rješenje *smisliti* ne odgovara značenju i kontekstu. Unatoč toj pogreški sustava za strojno prevođenje, smisao rečenice relativno je zadržan i proziran. Iz segmenata poput ovih vidi se zašto je jedna od glavnih zamjerki za neuronske sustave to da su prijevodi neprecizni i da ostavljaju dojam ukočenih i neprirodnih tekstova, što je potvrđeno i rezultatima ovog istraživanja, ali se i iz takvih prijevoda svejedno može iščitati većina značenja. Nepreciznost je uzrok teškoća i u Primjeru 23:

Primjer 23. *Sada postoje škole u kojima je uvedeno kontinuirano ocjenjivanje nastavnika od strane učenika.*

Referentni prijevod: U **međuvremenu** postoje škole u kojima je uvedeno kontinuirano vrednovanje nastavnika **koje provode učenici**.

Izvornik: *Es gibt **inzwischen** Schulen, an denen eine kontinuierliche Lehrerbewertung **durch die Schüler** eingeführt wurde.*

Prilog *sada* iz Primjera 23 ne odgovara sasvim njemačkom prilogu *inzwischen*: ako već nije odabran izravni ekvivalent *u međuvremenu*, kao što je u referentnom prijevodu, prilogu *sada*

trebalo bi dati neku dopunu da se prenese pragmatička funkcija, npr. *sada već postoje škole* (...). I to je anotator ubrojio kao jednu od leksičkih pogrešaka.

U istom primjeru anotator je označio još jednu pogrešku koja počiva na nepreciznosti, ali nju je smjestio u kategoriju „ostalo“. Anotator je alatu za strojno prevođenje zamjerio izraz *od strane učenika*, smatrajući ga neidiomatskim. S druge strane, treba uzeti u obzir da se konstrukcija *od strane nekoga*, premda doista nije u skladu s preporukama gramatičkih i pravopisnih priručnika za hrvatski jezik, ipak javlja i u brojnim konvencionalnim prijevodima, kao i tekstovima izvorno napisanim na hrvatskom jeziku, pogotovo u stručnom žargonu. Ovaj anotator to je mjesto legitimno označio, smatrajući ga pogreškom koju bi trebalo ispraviti i u konvencionalnom prijevodu, koju stoga nije htio propustiti označiti ni u strojnom prijevodu, ali to je također još jedan od primjera gdje se ne bi svi anotatori nužno lako složili. Sasvim je moguće da bi neki drugi anotator zanemario takvu konstrukciju jer se može argumentirati da je prisutna u jeziku.

Anotatori, dakle, mogu biti stroži i blaži, i njihovi pogledi na jezik snažno mogu varirati, toliko da bi se broj anotiranih pogrešaka mogao znatno razlikovati da je odabran drugi anotator, što samo još jednom naglašava koliko je doista njihova uloga u postupku analize pogrešaka i opravdava činjenicu da se analiza pogrešaka ne može propustiti laicima.

Zaključno se za IP5 (*Utječu li pogreške u leksičkoj točnosti bitnije na nerazumijevanje prijevoda od ostalih pogrešaka?*) može reći da leksičke pogreške u ovom istraživanju sačinjavaju gotovo polovicu svih zabilježenih pogrešaka, ali nisu izravno i samostalno utjecale na nerazumljivost teksta jer, uslijed sve veće točnosti unaprijeđenih strojnih sustava za prevođenje, te pogreške najčešće nisu bile toliko ozbiljne ili je njihovo značenje svejedno bilo relativno prozirno i lako dokučivo iz konteksta. Istražujući utjecaj leksičkih pogrešaka na percipiranu razumljivost, zaključuje se da je to pitanje od mnogo manjeg značaja u neuronskim strojnim prijevodima nego što je bio slučaj u statističkim prijevodima, a ukazuje se i potreba za redefiniranjem kategorije leksičkih pogrešaka. Grube leksičke pogreške kakve smo poznavali iz vremena statističkih alata za strojno prevođenje sada postoje samo u tragovima, a umjesto njih javljaju se neprecizna, nezgrapna i neidiomatska prijevodna rješenja.

Za budućnost istraživanja vrednovanja strojnog prevođenja stoga treba razmisliti o redefiniranju kategorija koje su već desetljećima etablirane unutar discipline strojnog prevođenja. O potrebi personaliziranja kategorizacije pogrešaka ovisno o jezičnom paru ili namjeni analize pogrešaka već se pisalo u mnogim radovima: Popović (2018: 25) priključuje

se zaključku da najbolja praksa provođenja analize pogrešaka uključuje veliku generaliziranu taksonomiju iz koje se po potrebi izvlači prikladan manji set pogrešaka za određenu analizu, kao što to Klubička, Toral i Sánchez-Cartagena (2018: 8) čine prilagođavajući taksonomiju MQM-a za slavenske jezike. U tom se istraživanju unutar kriterija tečnosti uklonila kategorija *tipografija*, i na njezino se mjesto stavio (jezični) *registar*, jer je preliminarno istraživanje pokazalo određene devijacije od standardnog jezika koje bi korisno bilo izdvojiti u zasebnu kategoriju. Temeljeći se na prijašnjim istraživanja i u ovom je dokazana korisnost kategorije „ostalo“ u prevođenju na hrvatski jezik, premda je ta potreba u kombinaciji s njemačkim jezikom i s neuronskim sustavom bila mnogo manje izražena nego što je to bio slučaj sa statističkim prijevodom sa švedskog na hrvatski kada je gotovo 2 % segmenata bilo posve nerazumljivo (Ljubas 2017: 44). Taj je fenomen pokazatelj da kvaliteta strojnih prijevoda neprestano raste. Izlazni podaci statističkih sustava koji su vrvjeli od velikog broja najrazličitijih pogrešaka najzad su stvar prošlosti, te je za neuronske sustave potrebno redefinirati kategorije pogrešaka da bi se iz njih izvukli nijansiraniji zaključci. Međutim, to ne znači da je u analizi pogrešaka potrebno imati više potkategorija, već da i dalje treba zadržati šest-sedam kategorija, samo što bi one sada trebale biti uže definirane. Kao što su se u statističkim prijevodima zamijetile poteškoće s redosljedom riječi u rečenici, zbog čega se ta potkategorija sintaktičkih pogrešaka izdvojila kao zasebna kategorija, tako treba pratiti ishode recentnijih istraživanja na neuronskim sustavima i sukladno tome ažurirati kategorije. Tako, primjerice, rezultati ovog istraživanja daju naslutiti da su nedovoljno precizna ili nedovoljno idiomatski izražena rješenja, kao u Primjerima 21-23, novi izazov strojnih prijevoda budućnosti. Promjene koje je potrebno uvoditi u metode vrednovanja dodatno se tematiziraju u sljedećim dvama potpoglavljima.

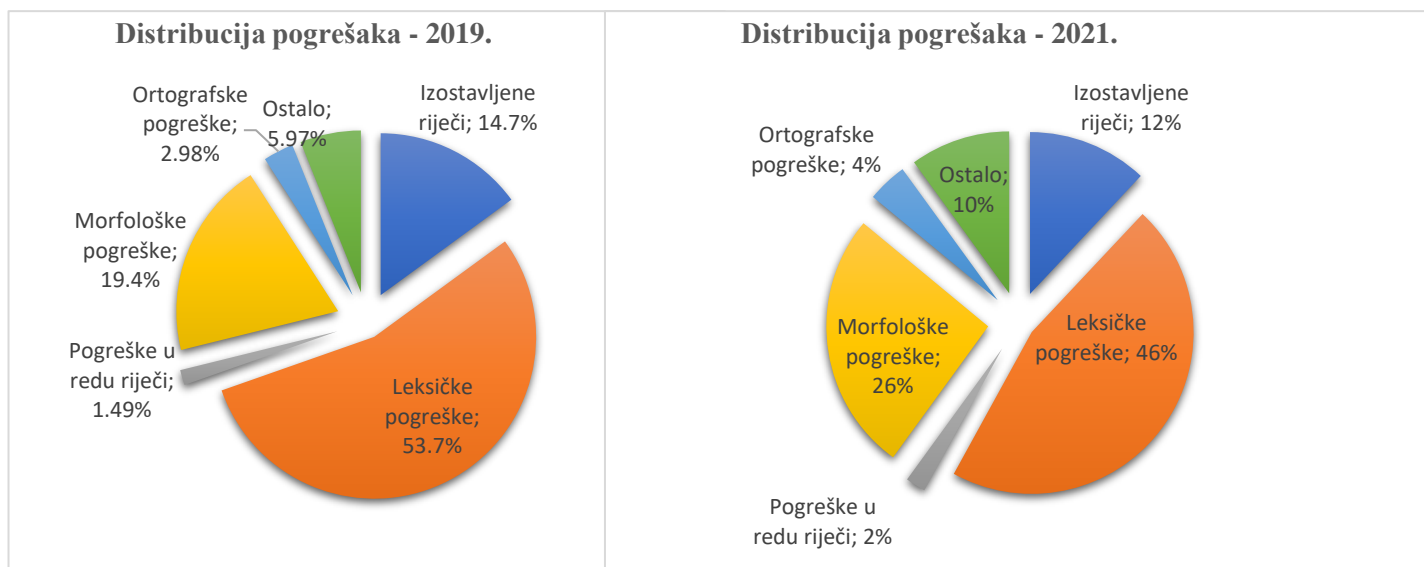
5.4 Dijakronijska usporedba rezultata po pojedinim metodama vrednovanja (IP6, rezultati i analiza)

Rezultati ovog istraživanja ukazuju na to da se navedene metode vrednovanja teže nose s poboljšanjem jezične kvalitete strojnih prijevoda odnosno da podaci dobiveni tim metodama nisu jednako relevantni kao u prošlosti. U prošlom potpoglavlju predstavljena je analiza pogrešaka na strojnom prijevodu koji je generiran 2021. godine, ali isti je tekst strojno preveden i dvije godine prije (2019.) za predistraživanje, i analizu je proveo isti anotator. U prijevodu je

tada zabilježeno šezdeset i osam pogrešaka od ukupno četriristo devedeset riječi u prevedenom tekstu, a od toga:

- deset izostavljenih riječi (četiri više nego 2021.)
- trideset i šest pogrešnih leksičkih odabira (trinaest više nego 2021.)
- trinaest morfosintaktičkih pogrešaka (isto kao i 2021.)
- četiri pogreške u ortografiji (dvije više nego 2021.);
- jednu pogrešku u redosljedu riječi (i 2021. zabilježena je jedna pogreška u redosljedu riječi, ali ne ista; pogreška se javlja u drugom segmentu);
- četiri ostale pogreške (jedna manje u odnosu na 2021.).

Ta je distribucija usporedno prikazana na Grafu 8 i Grafu 9:



Graf 8: Anotirane pogreške, 2019.

Graf 9: Anotirane pogreške, 2021.

Odmak od dvije godine utjecao je na veliko poboljšanje u jezičnoj kvaliteti teksta, osobito u leksičkoj točnosti izlaznih podataka. Primjer 13 analiziran je u poglavlju 5.3.1 kao primjer ortografske pogreške, no to je ujedno i segment koji sadržava najviše pogrešaka. Ovaj put navedeni su prvo izlazni podaci iz 2019. godine, a zatim se za usporedbu ponavljaju podaci iz 2021. godine:

Primjer 13. (2019.): "Izraz" *Ja sam učitelj* "uvijek \emptyset se čini kao **ispovijed**, kao **obveza prema** nečemu negativnom, poput **obveze prema** poroku **pušenje** ili još **gore**.

Primjer 13. (2021.): Rečenica "Ja sam učitelj" uvijek mi se čini kao priznanje, kao priznanje nečemu negativnom, otprilike slično priznanju poroku poput pušenja ili još gore.

Referentni prijevod: Rečenica „Ja sam nastavnica“ uvijek mi zvuči kao neko priznanje, kao da priznajem nešto negativno, otprilike kao da imam neki porok, recimo da pušim ili što još gore.

Izvornik: *Der Satz „Ich bin Lehrerin“ kommt mir stets wie ein Bekenntnis vor, wie ein Bekenntnis zu etwas Negativem, ungefähr ähnlich dem Bekenntnis zu einem Laster wie dem Rauchen oder Schlimmerem.*

Premda se u ovom istraživanju od anotatora nije tražilo da pogreškama pridruži određeni koeficijent „težine“ pogreške, ipak se, uspoređujući izlazne podatke generirane u različitim vremenskim razdobljima, uočava određena razlika u njihovoj ozbiljnosti. Već i ortografska pogreška koja se javlja u oba slučaja, 2019. je godine ozbiljnija nego što je bila u rezultatima iz 2021. godine (kada se ticala tek konvencije mjesta postavljanja znakova navoda u hrvatskom jeziku i čitatelj ju je zapravo sasvim lako mogao zanemariti). Kao prvo, pod znakove je navoda stavljen potpuno pogrešan dio rečenice ("Izraz" odnosno *Der Satz* iz izvornika, umjesto *Ich bin Lehrerin*), kao drugo ni ti znakovi ne prate konvencije hrvatskog jezika, a kao treće javlja se i treći znak navoda koji u izvorniku ne postoji, a u izlaznim podacima otvara navod koji se nikada ne zatvara. Tako je od jednog elementa u znaku navoda došlo do triju ortografskih pogrešaka jer se svaki ovaj znak nalazi na pogrešnom mjestu, i to je nešto što čitatelje svakako može zbuniti.

Nadalje, dolazi do pogreške u kategoriji izostavljenih riječi: u prijevodu njemačkog izraz *kommt mir vor* izgubila se zamjenica *mir* pa se ekvivalent našao u nepersonalnom obliku *čini se* umjesto *čini mi se* (ili *uvijek mi zvuči*, kao što je u referentnom prijevodu). Čitatelj u tom slučaju mora pretpostaviti da je to nekakvo opće značenje, a ne mišljenje autorice teksta.

Ostale su pogreške u Primjeru 13 leksičke – 2021. od njih je zadržana samo jedna, i to ona koja se ponavlja u cijelom tekstu: *Lehrerin* je prevedeno kao *učitelj* umjesto *nastavnica*¹⁷, ali

¹⁷ Sustavi za strojno prevođenje u zadnje su vrijeme često kritizirani zbog seksizma ili perpetuiranja ljudskog seksizma u prijevodima. I u ovom je primjeru uočeno da je njemačka imenica *Lehrerin*, koja sadrži nastavak za ženski gramatički rod, prevedena u muškom gramatičkom rodu, vjerojatno pod utjecajem opsežnog korpusa gdje

2019. se javljaju i mnogi drugi pogrešni leksički odabiri: *izraz* umjesto *rečenica* za njemačko *der Satz*; rješenje *ispovijed* nikako ne odgovara ovom kontekstu za riječ *Bekanntnis*; a rečenica završava i vrlo nejasnim dijelom *poput obveze prema poroku pušenje ili još gore*, koji je k tome još i morfosintaktički neusklađen. Osoba kojoj bi bio posvećen zadatak redakture ove rečenice (ili općenito izlaznih podataka iz 2019. godine) potencijalno bi brže i s manjim kognitivnim i tehnološkim izazovima obavila posao ako bi rečenicu prevela ispočetka, potpuno zanemarujući strojni prijevod.

Da je prijevod nerazumljivim smatralo i četrdeset i petero ispitanika preliminarnog istraživanja, govori nam prosječna ocjena razumljivosti: 2019. godine isti strojno prevedeni tekst dobiva ocjenu razumljivosti 2,95, što je osjetno niže od ocjene razumljivosti strojnog prijevoda iz 2021. godine, a koja je iznosila 3,6 kod ispitanika koji su vrednovali tekst kao cjelinu (u istim uvjetima kao ispitanici predistraživanja), a svakako onda još mnogo manja od ocjene koju su tekstu dodijelili ispitanici koji su vrednovali raščlanjene segmente, a ona je iznosila 4,1. U Tablici 2 prikazane su ocjene razumljivosti svih ispitanika predistraživanja (kako se ispitanici predistraživanja razlikuju od ispitanika glavnog istraživanja, oni su označeni malim početnim slovom „i“):

Tablica 2: Ocjene razumljivosti, 2019.

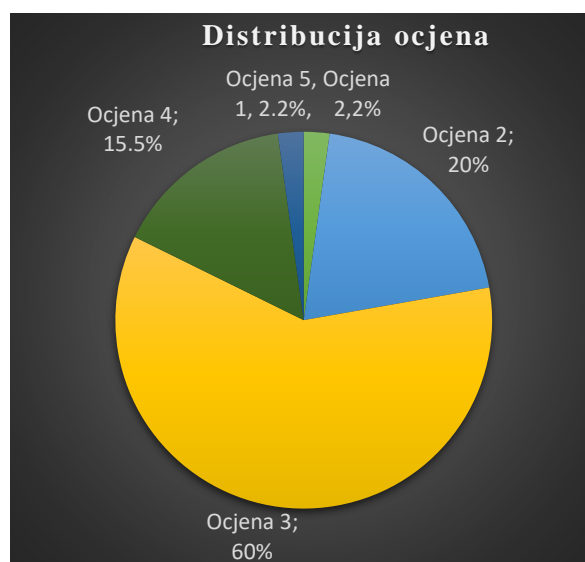
Ispitanik	Ocjena cjelokupnog segmenta
i1	3
i2	3
i3	3
i4	2
i5	3
i6	3
i7	4
i8	2
i9	4

se muški gramatički rod rabi kao neutralan oblik za biološke ljudske spolove i sociološke rodove. Ipak, ne treba zaboraviti da sustavi za strojno prevođenje samo zrcale naučene obrasce koji se javljaju u tekstovima na kojima su trenirani. Za pregled dosadašnjih radova na ovu temu v. Savoldi i sur. 2021.

i10	4
i11	3
i12	3
i13	5
i14	3
i15	2
i16	3
i17	4
i18	2
i19	3
i20	3
i21	3
i22	3
i23	3
i24	3
i25	3
i26	3
i27	2
i28	2
i29	3
i30	3
i31	2
i32	3
i33	3
i34	1
i35	2
i36	3
i37	3
i38	4
i39	3

i40	3
i41	4
i42	4
i43	3
i44	3
i45	2
Ukupna ocjena:	2,95

U ovom je predistraživanju (Ljubas 2019) donekle postignuto slaganje među ispitanicima: dvadeset i sedmero je ispitanika (60 %) izlazne podatke ocijenilo ocjenom 3. U svjetlu činjenice da je 2021. postignuto slaganje od tek 40 % u skupini ispitanika koji su vrednovali cjelokupni tekst, može se reći da su ovi rezultati vjerodostojniji. Ostale ocjene prilično jasno slijede normalnu distribuciju: ocjenu 4 odabralo je sedmero ispitanika (15,5 %), a ocjenu 2 devetero ispitanika (20 %). Samo po jedan ispitanik (2,2 %) prijevod je ocijenilo najnižom ocjenom 1 odnosno najvišom ocjenom 5 (v. Graf 10 za distribuciju).



Graf 10: Distribucija ocjena, 2019.

Premda je u preliminarnom istraživanju sudjelovalo manje ispitanika, rezultati ukazuju na to da se povećanjem jezične kvalitete otežava postupak subjektivne procjene razumljivosti teksta. To se može objasniti činjenicom da se u bolje prevedenim tekstovima ispitanici imaju bolju priliku

usredotočiti na suptilnije nijanse – za koje će neki ispitanici na koncu ipak zaključiti da nisu bitne u široj slici, a drugi će na temelju njih umanjiti svoje ocjene, čime nastaje veći nesrazmjer.

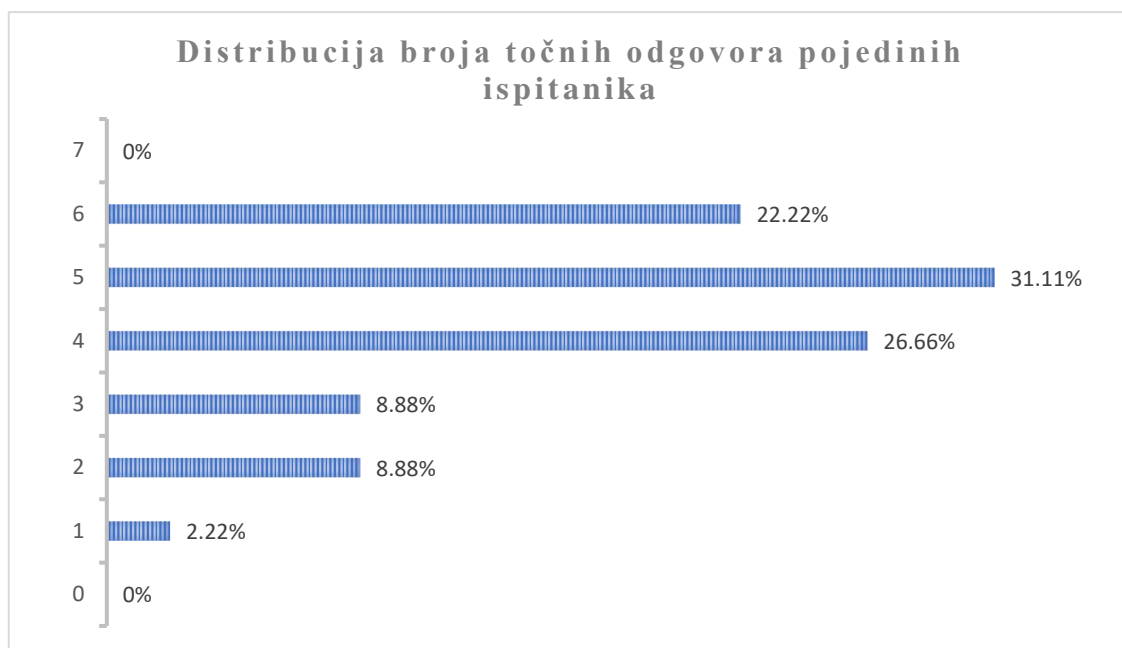
IP6 glasilo je *Postiže li se smanjenjem broja jezičnih pogrešaka u izlaznim podacima veća dosljednost u rezultatima vrednovanja metodama subjektivne procjene i testa razumljivanja?*

Iz rezultata se može zaključiti da se smanjenjem broja jezičnih pogrešaka u izlaznim podacima ne postiže veća dosljednost u rezultatima vrednovanja subjektivnom procjenom razumljivosti, štoviše, događa se upravo suprotno – slaganje među ispitanicima sve više opada što je jezična kvaliteta izlaznih podataka bolja, jer se u slučajevima boljih rezultata strojnih prijevoda pažnja ispitanika usmjerava na detalje i nijanse, čiju važnost svi različito tumače. Može se pretpostaviti da su ispitanici u slučaju što boljih strojnih prijevoda skloniji uspoređivati ih s ljudskim prijevodima i snažnije im zamjerati one pogreške koje bi zamjerali i ljudskom prevoditelju, no tako ne razmišljaju svi ispitanici, pa se tu javlja nesrazmjer u njihovim subjektivnim procjenama. Kod grubljih je pogrešaka pažnja većine ispitanika toliko usmjerena na njih da se paradoksalno postiže veće slaganje. Također se može argumentirati da su danas neki ispitanici više izloženi strojnim prijevodima u svakodnevnom životu, pa su naviknuti na vrstu pogrešaka kakve se u njima nalaze, dok se drugima one mogu činiti posve stranima i nelogičnima.

A kako se smanjenje broja jezičnih pogrešaka manifestira kod provedbe metode testa razumljivosti? U drugoj tablici Priloga D prikazana je distribucija točnih i netočnih odgovora svih ispitanika preliminarnog istraživanja iz koje je razvidno da:

- nema ispitanika koji nisu točno odgovorili ni na jedno pitanje;
- jedan je ispitanik točno odgovorio na jedno pitanje;
- četvero je ispitanika točno odgovorilo na dva pitanja;
- četvero je ispitanika točno odgovorilo na tri pitanja;
- dvanaestero je ispitanika točno odgovorilo na četiri pitanja;
- četrnaestero je ispitanika točno odgovorilo na pet pitanja;
- desetero je ispitanika točno odgovorilo na šest pitanja;
- nema ispitanika koji su točno odgovorili na svih sedam postavljenih pitanja.

U Grafu 11 ova je distribucija prikazana u postocima:



Graf 11: Distribucija broja točnih odgovora – preliminarno istraživanje, 2019.

Razvidno je da je, jednako kao i 2021. gotovo 80 % ispitanika točno odgovorilo na više od pola pitanja, samo se distribucija ponešto razlikuje:

- 2019. godine nitko nije točno odgovorio na sva pitanja, za razliku od 8,75 % ispitanika 2021. godine (v. i Graf 5 u potpoglavlju 5.2.1 za usporedbu);
- 2019. godine 22,22 % ispitanika točno je odgovorilo na šest pitanja, za razliku od 18,75 % ispitanika 2021. godine;
- 2019. godine 31,11 % ispitanika točno je odgovorilo na pet pitanja, za razliku od 28,75 % ispitanika 2021. godine;
- 2019. godine 26,66 % ispitanika točno je odgovorilo na četiri pitanja, za razliku od 22,5 % ispitanika 2021. godine;
- 2019. godine 8,88 % ispitanika točno je odgovorilo na tri kao i na dva pitanja, za razliku od neujednačenije distribucije 2021. godine, kada je 14,3 % ispitanika odgovorilo na tri pitanja, a 5 % na dva pitanja;
- 2019. godine 2,22 % ispitanika točno je odgovorilo na jedno pitanje, za razliku od 0,62 % 2021. godine;
- 2019. nema ispitanika s nijednim točnim odgovorom, što se 2021. godine promijenilo (1,25 %).

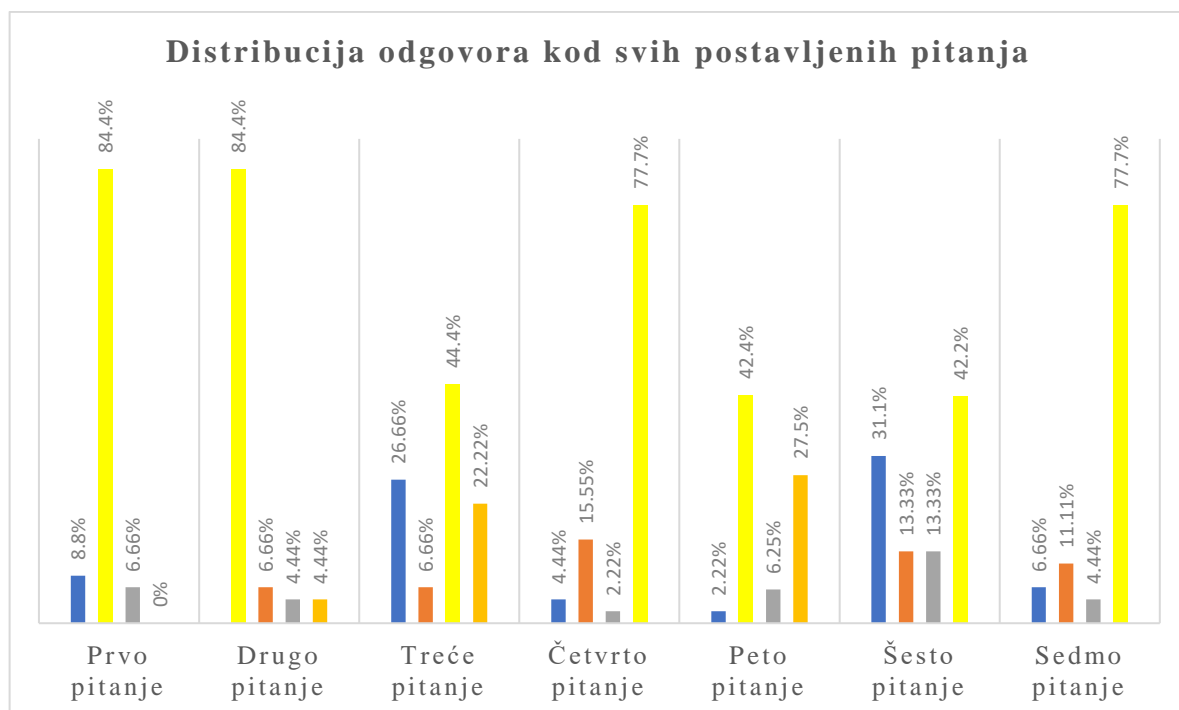
Valja još usporediti i distribuciju točnih odgovora po pojedinim pitanjima:

- 2019. godine na prvo i drugo pitanje točno je odgovorilo trideset i osmero ispitanika, što čini 84,44 %, i to je tek nešto niži rezultat nego 2021. godine kada je na prvo pitanje točno odgovorilo 87,5 % ispitanika, a na drugo 91,8 % ispitanika;
- 2019. godine na treće pitanje točno je odgovorilo dvadesetero ispitanika, što čini 44,4 %, ponovno tek nešto manje od 2021. godine kada je na to pitanje točno odgovorilo 48,1 % ispitanika;
- 2019. godine na četvrto i sedmo pitanje točno je odgovorilo trideset i petero ispitanika, što čini 77,7 % i vrlo je slično rezultatu iz 2021. godine, kada je na četvrto pitanje točno odgovorilo 78,1 % ispitanika, a na sedmo 79,3 % ispitanika.

Ispitanici su, osim s trećim pitanjem, 2021. godine najviše poteškoća imali s petim i šestim pitanjem, no 2019. godine to se nije tako dogodilo:

- 2019. godine najveći je broj ispitanika na peto pitanje odgovorio točno (42,2 % ispitanika, njih devetnaestero), a odgovor C odabralo je tek 6,25 % ispitanika. Za usporedbu, 2021. godine na peto je pitanje više ispitanika odabralo netočan odgovor C (36,25 %) u odnosu na točan odgovor B (30,6 %). Objе je godine netočan odgovor D odabralo 27,5 % ispitanika;
- 2019. godine na šesto je pitanje više ispitanika odabralo netočan odgovor D (njih 42,2 %, devetnaestero ispitanika), nego točan odgovor A koji je odabralo 31,1 % ispitanika, njih četrnaestero. Za usporedbu, 2021. godine na šesto je pitanje točno odgovorilo 46,25 % ispitanika, a također je 41,25 % ispitanika odabralo netočan odgovor D.

Ta je distribucija vidljiva na Grafu 12:



Graf 12: Postotna uspješnost ispitanika na svim pitanjima

Što nam sve ove brojke govore? Rezultati metode testa razumljivosti 2019. i 2021. vrlo su slični – blaga odstupanja u broju točnih odgovora mogu se pripisati tome što je tekst 2019. pod utjecajem većeg broja pogrešaka doista mjestimično mogao biti teže razumljiv. Dijakronijska dosljednost u rezultatima svjedoči o tome da su rezultati dobiveni metodom testa razumijevanja objektivniji od onih dobivenih drugim metodama – istraživač ih doista može smatrati pouzdanima. Odgovor na IP6 u slučaju metode testa razumljivosti glasi da se smanjenjem broja jezičnih pogrešaka ne postiže nužno veća dosljednost u rezultatima vrednovanja, ali dijakronijski rezultati snažno ukazuju na to da su rezultati dobiveni metodom testa razumijevanja već dovoljno objektivni i pouzdani, a u svakom su slučaju objektivniji nego oni dobiveni metodom subjektivne procjene.

Na koncu još valja proučiti je li niska ocjena razumljivosti u predistraživanju 2019. povezana s tadašnjim stavovima ispitanika. Anketni obrazac o stavovima ispitanika pokazao je da je i 2019. godine 71,1 % ispitanika pokazalo pretežno pozitivne stavove prema strojnom prevođenju, 20 % ispitanika imalo je neutralan stav, a 8,8 % ispitanika imalo je pretežno negativne stavove. Rezultati su vrlo slični onima iz 2021. godine kada je 77,5 % ispitanika imalo pozitivan stav, 19,37 % neutralan stav, a samo 3,1 % pretežno negativan stav prema strojnom prevođenju. Sitna odstupanja ponovno su na strani alata za strojno prevođenje,

odnosno dijakronijski rezultati analize stavova govore u prilog tome da su ljudi i dalje sve pozitivnije nastrojeni prema njima.

To ujedno znači i da se relativno niska ocjena subjektivne razumljivosti iz 2019. ne može dovesti u jasnu svezu s negativnim stavovima, odnosno ne bi se moglo reći da su ispitanici zbog svojih prijašnjih mišljenja prijevodu davali slabije ocjene, već se objašnjenje mora potražiti ili doista u jezičnim pogreškama kojih je bilo više ili u tome što je narav subjektivne metode vrlo proizvoljna i da podaci dobiveni njome nisu dovoljno vjerodostojni. Činjenica da su na testu razumijevanja i 2019. rezultati bili vrlo dobri također je u skladu s komentarima koje su ispitanici predistraživanja svojevolumno dopisali u otvorenom polju na kraju istraživanja:

i32, plesačica: *Ako se prevodi jedna riječ ili jednostavna rečenica, onda je razumljivo, ali ako se prevodi kompliciraniji tekst, prijevod je pun grešaka ili nije u potpunosti razumljiv.*

i39, nastavnica geografije i povijesti: *Uglavnom su dobri kada se prevodi samo jedna riječ, ali kada se radi o duljem tekstu često se izgubi smisao.*

i40, pravnica: *Kada sam se koristila alatima za strojno prevođenje nikada nisam bila u potpunosti zadovoljna, budući da je prijevod uvijek bio nejasno i da se gubio cijeli kontekst.*

Ispitanici i32 i i39 imaju pretežno pozitivne stavove prema strojnom prevođenju, a i40 neutralan stav, a sve su tri ispitanice razumljivost prijevoda ocijenile ocjenom 3, što je i prosjek kada se uzmu u obzir ocjene drugih ispitanika. Iz ovih se stavova može iščitati da su ispitanici katkad i previše svjesni pogrešaka koje programi za strojno prevođenje rade i da smatraju da se zbog njih „izgubi smisao“ ili „cijeli kontekst“, što svjedoči o tome da zapravo ni sami nisu svjesni koliko toga i unatoč pogreškama jesu u stanju razumjeti. Ispitanica i40 navodi da nikada nije „bila u potpunosti zadovoljna“ strojnim prijevodima, ali i na ovom je testu točno odgovorila na većinu pitanja. I kod i32 iščitava se potreba za razumijevanjem teksta u *potpunosti*. Na temelju stavova ispitanika može se zaključiti da jezična ispravnost teksta znatno više utječe na subjektivnu procjenu nego na objektivno razumijevanje teksta, što je još jedan argument u

prilog metodi vrednovanja testom razumijevanja kao načinu objektivnog prikupljanja podataka o stvarnoj razumljivosti strojnih prijevoda.

5.5 Usporedba istraženih ljudskih metoda vrednovanja (završna diskusija)

U tumačenju rezultata i raspravi o istraživačkim pitanjima ne treba zaboraviti s kojim su ciljem ona postavljena. U ovom su se istraživanju mijenjali parametri metoda vrednovanja ne bi li se otkrilo je li moguće stvoriti optimalne uvjete za donošenje vjerodostojnih sudova o kvaliteti prijevoda. Krajnje je pitanje zapravo koju od ovih metoda istraživač strojnog prevođenja može odabrati s povjerenjem kada želi provjeriti kvalitetu izlaznih podataka nekog teksta, s polazišnim kriterijem razumljivosti iz perspektive korisnika koji nisu profesionalni prevoditelji, već se strojnim prijevodima služe u privatne svrhe za razumijevanje tekstova koji nisu napisani na njihovu materinskom ili nekom stranom jeziku koji govore.

Metodom subjektivne procjene pokušala se iskoristiti subjektivna narav tog postupka vrednovanja, ali kao najveći se izazov brzo pokazalo to što je među ispitanicima doista teško postići ikakvu relevantnu razinu slaganja. U ovom istraživanju skupinu ispitanika činili su neprevoditelji koji su eventualna iskustva s alatima za strojno prevođenje stekli u privatnom okruženju ili se njime služe poslovno, ali ne kao jezičari i prevoditelji, već za druge namjene. Uočeno je da su njihovi stavovi o strojnom prevođenju poprilično homogeno pozitivni. Suprotno očekivanjima koja su se temeljila na prijašnjim istraživanjima, čak je 77,5 % ispitanika iskazalo pretežno pozitivne stavove prema strojnom prevođenju, a ispitanici su također pokazali svijest o tome koliko su okruženi strojnim prijevodima, kao i da se njima aktivno služe. Upravo je zbog toga važno vršiti istraživanja o strojnom prevođenju i s neprevoditeljima jer su svrhe u koje se ljudi njima služe doista raznovrsne.

Sami su ispitanici zamijetili pomake koji se u strojnom prevođenju događaju proteklih godina kod javno dostupnih alata za strojno prevođenje, prvenstveno Google Prevoditelja, što koincidira s njegovim prelaskom na neuronski model prevođenja. U istraživanju je sudjelovalo sto šezdeset ispitanika, koji su izlazne podatke vrednovali na ljestvici s pet intuitivno razumljivih ocjena s pridruženim objašnjenjima od „nerazumljivo“ do „potpuno razumljivo“. U skladu s recentnijim istraživanjima (v. Castilho i sur. 2018), s porastom broja ispitanika nije rasla razina slaganja, već je opadala, što je i dalje primarni izazov ove metode. Promjena parametra jedinice vrednovanja također nije bitnije utjecala na postizanje slaganja među

ispitanicima, što znači da nije bilo moguće pružiti odgovor na istraživačko pitanje o tome postiže li se veće slaganje kada se vrednuju raščlanjeni segmenti ili cjelokupan tekst.

Međutim, ovo je istraživanje otkrilo da se metode vrednovanja s porastom kvalitete strojnih prijevoda nalaze pred novim izazovima, pa to treba uzeti u obzir u budućim istraživanjima vrednovanja strojnog prevođenja. U neuronskom strojnom prevođenju nailazimo na sve manje leksičkih i morfosintaktičkih pogrešaka kojima su u statističkim sustavima tekstovi znali vrvjeti (v. npr. Ljubas 2018), što znači da fokus ispitanika u novoj eri treba usmjeriti na istančanije vrste pogrešaka. Te se pogreške mogu odnositi na nadrečenične elemente ili se u rečenicama mogu uočiti samo onda kada ih se promatra u njihovu širem kontekstu. Kada su vrednovatelju dijelovi teksta prikazani raščlanjeno i nasumično, kao što je dosad bilo uobičajeno, on se na takve pogreške nema priliku osvrnuti. To je osobito važno upravo kada se kao kriterij vrednovanja odabere razumljivost, umjesto, primjerice, točnosti ili tečnosti, jer tekstovi pisani na prirodnim jezicima uvijek sadrže deiktičke elemente i svi se dijelovi i u izvorniku trebaju čitati unutar konteksta da bi bili razumljivi, stoga se zaključno može preporučiti da se strojni prijevodi u cjelini i vrednuju.

Nedostaci metode subjektivne procjene u ovom su se istraživanju pokazali većima od njezinih prednosti – brzine pripreme i provedbe te relativne lakoće prikupljanja podataka. Stavovi ispitanika pokazuju da se znaju služiti strojnim prijevodima u svrhe za koje su im *dovoljno dobri* te da imaju svakodnevnu korist od njih, ali ovo je istraživanje pokazalo da se metodom subjektivne procjene ne može doista smisleno iščitati koliko su zapravo informacija ispitanici izvukli iz nekog teksta. Podaci dobiveni ovom metodom – brojčane ocjene – u tom smislu podsjećaju na podatke koji se dobivaju automatskim metrikama za vrednovanje prijevoda: nije jasno kako tumačiti ocjenu svakog pojedinog ispitanika kada su one tako nasumične. Teza „zlatnog standarda“ ljudske subjektivne procjene poljuljana je manjkom slaganja među ispitanicima, i zapravo ju je potrebno drukčije tumačiti: zapažanja ljudi, koji doista jesu jedini koji smisleno mogu izraziti jesu li im neki podaci razumljivi, bez općeg se slaganja i usklađenosti u mišljenjima samo individualno mogu i tumačiti – jednom čovjeku u jednom je trenutku u vremenu određeni strojni prijevod razumljiv do određene razine. Iako i to govori nešto, u istraživačke svrhe ti podaci nisu dovoljno relevantni.

Testovi razumijevanja tome su jasna opreka: podaci dobiveni njima pružaju mnogo jasniji uvid u to koji su dijelovi teksta dovoljno razumljivo prevedeni da bi se iz njih izvukle ključne informacije te koji su dijelovi teksta bolje prevedeni od drugih. Također, ako su pitanja vješto

osmišljena, ispitanici moraju pokazati razumijevanje cjelokupnog teksta, pa pozitivni rezultati kod većine ispitanika jednoznačno ukazuju na kvalitetno preveden tekst. Još je jedna prednost metode testa razumijevanja: za razliku od metode subjektivne procjene, moguće je neposredno promatrati utjecaj na rezultate onih čimbenika koji sa samim strojnim prijevodom nisu povezani – kao što su stavovi ispitanika ili kognitivni naponi koje su uložili u vrednovanje.

Premda istraživač i dalje kvalitetu prijevoda prosuđuje samo po krajnjim podacima – broju točnih odgovora koje su ispitanici ponudili – s pomoću dodatnih pitanja koja su uvedena u ovom istraživanju o sigurnosti odabira odgovora, mjerenju vremena potrebnog za rješavanje testa i definiranju svojih stavova, istraživač ima mogućnost pratiti suodnos svih tih faktora. Ispitanici su k tome nerijetko i sami nesvjesni koliko su informacija doista izvukli iz prijevoda jer ga tijekom čitanja uspoređuju s tekstovima pisanim na prirodnom jeziku ili konvencionalnim prijevodima koji obično ne sadržavaju pogreške kakve se javljaju kod strojnih prijevoda. Stoga njihove subjektivne procjene doista iskreno mogu odražavati ono što misle o razumljivosti prijevoda, ali na testu razumijevanja ipak se može pokazati nešto drugo, kao što je i bio čest slučaj u ovom istraživanju. S druge strane, neki ispitanici također mogu misliti da pri procjeni razumljivosti ili izražavanju stavova njihovi odgovori moraju biti „ispravni“ ili oni za koje misle da su poželjni, ali svi ti prirodni ljudski mehanizmi gube na svojoj važnosti kada je metoda objektivna.

Saznanje da je 80 % ispitanika točno odgovorilo na više od pola postavljenih pitanja, od toga 8,75 % na sva pitanja testa razumijevanja, rezultat je koji govori više o razumljivosti prijevoda nego samo jedna ocjena. Ispitivač k tome vidi i neke nijanse: dio teksta u kojemu se krije odgovor na pitanja na koje je čak 91,8 % ispitanika odgovorilo točno razumljivo je preveden, a oni dijelovi teksta gdje je manje ispitanika odgovorilo točno sadrže neka izazovnija mjesta. Kada bi subjektivne ocjene segmenata korelirale s tim rezultatima, odnosno da su visoke ocjene razumljivosti dodijeljene segmentima u kojima su sadržani odgovori na pitanja na koja je većina ispitanika odgovorila točno, a niske ocjene onima gdje su ispitanici imali poteškoća, dokazivalo bi to da se i subjektivnom procjenom može doći do jednako objektivnih rezultata, i to bržom metodom. To ipak nije slučaj, pa prednost treba dati metodi testa razumijevanja.

Usto, u pripremi ovog istraživanja pokazalo se da glavni nedostatak ove metode, zamornost pripreme testa, nije nepremostiv – u međuvremenu osim velikog broja javno dostupnih standardiziranih testova iz međunarodnih ispita za mnoge jezike, istraživači vrlo lako mogu doći i do brojnih usporednih korpusa s izvornicima i referentnim prijevodima. Alternativno se

od ispitanika može tražiti da, umjesto da odgovaraju na pitanja, izvedu neki zadatak čije su upute strojno prevedene, ovisno o vrsti teksta. A čak i ako se test razumijevanja treba pripremiti ispočetka, rezultati ovog istraživanja u svakom slučaju govore u prilog tome da se dodatan trud uloženi u provedbu ove metode isplati zbog kvalitetnih podataka o razumljivosti koji se njome postignu.

Budući da se izlazni podaci alata za strojno prevođenje neprestano poboljšavaju i da prijevodi postaju sve razumljiviji, između ostalog je vrlo važno da metode vrednovanja i u budućnosti mogu ostati mjerodavne. Dijakronijskom usporedbom metode subjektivne procjene utvrđeno je da je postignuta određena dosljednost među ispitanicima dok su izlazni podaci bili lošiji, što objašnjava njezinu razmjerno čestu uporabu u prošlosti vrednovanja, ali ukazuje i na to da mogućnosti primjene te metode drastično opadaju. S druge strane, s metodom testa razumijevanja postignuti su vrlo usporedivi rezultati u razmaku od dvije godine, što govori da ta metoda izdržava test vremena. Metoda testa razumijevanja, dakle, u dijakronijskom aspektu pokazuje veću izdržljivost, njezini su rezultati vjerodostojni i objektivni u okvirima naravi vrednovanja kao takvog. Također i broj jezičnih pogrešaka manje utječe na njezine rezultate nego u slučaju metode subjektivne procjene kada se vrednuje kriterij razumljivosti.

Analizom pogrešaka cjelokupni se proces vrednovanja predaje u ruke jednom anotatoru. U ovom istraživanju odabran je jezični stručnjak, obrazovani lingvist i prevoditelj. Kategorije pogrešaka naslanjale su se na one koje predlaže metrika MQM i na preliminarno istraživanje. Nakon preliminarnog istraživanja zadržane su samo one kategorije pogrešaka koje su pronađene u tom prijevodu, a ostale su uklonjene iz klasifikacije kako prevelik broj kategorija ne bi zamarao anotatora. Čak i u tim uvjetima, koji su najbliže idealnim što se moglo postići, rezultati analize pogrešaka i dalje su samo subjektivna opažanja jedne osobe. Istraživanje bi se moglo provesti i s više anotatora ili se od anotatora može zahtijevati da u skupini postignu slaganje oko analize pogrešaka, ali rezultati ostvareni kompromisom ne odražavaju nužno stvarno mišljenje pojedinih anotatora. Iz dosadašnjih je istraživanja također poznato da s porastom broja anotatora opada razina slaganja (v. Bayerl i Paul 2011).

Analiza pogrešaka također je još jedna od metoda kojoj opada značaj kako su izlazni podaci bolji – u ovom se istraživanju pokazalo kako su dvije godine odmaka utjecale na veliku razliku u broju pogrešaka, napose u leksičkim pogreškama, ali ono što se nije nužno pokazalo klasifikacijom jest ukupni dojam anotatora o tekstu ili koliko bi zahtjevno bilo preurediti tekst ako bi se objavljivao. U izlaznim je podacima iz 2021. godine zabilježeno manje pogrešaka, ali

anotator je u nestrukturiranom razgovoru obrazložio da mu se prijevod čini umjetan, plošan i doslovan, što je ponovno posljedica neuronskog strojnog prevođenja. Rezultati su u jednu ruku bolji, ali javljaju se i nove poteškoće koje se ne mogu nužno obuhvatiti postojećim kategorijama. To znači da kategorije pogrešaka zastarijevaju, iako je mnogo istraživanja posvećeno pravilnoj kategorizaciji pogrešaka i moglo bi se reći da čim se klasifikacija pogrešaka usavrši, dolaze nove promjene koje zahtijevaju drukčiji pristup.

Napose su lingvisti zainteresirani za analize pogrešaka jer se rezultati vrednovanja izražavaju u kategorijama s kojima oni procjenjuju i kvalitetu konvencionalnih prijevoda i jer su rezultati vrednovanja analizom pogrešaka izraženi u njima jasnim i poznatim kategorijama. Pa ipak, ako su u tekstu ovog istraživanja zabilježene dvadeset i tri leksičke pogreške i trinaest morfosintaktičkih pogrešaka, to ništa ne govori o tome koliko je tekst razumljiv čitatelju, odnosno metoda nije nužno prilagođena kriteriju vrednovanja. Ne znamo koliko je pogrešaka previše, ne znamo jesu li neke pogreške narušile percepciju razumljivosti ili stvarnu razumljivost više od drugih, samo dobivamo uvid u broj i narav pogrešaka. Kako je analiza pogrešaka također vrlo mukotrpa metoda, pitanje je opravdavaju li dobiveni podaci provedbu tog postupka. Stoga je i u praksi analiza pogrešaka u mnogim slučajevima samo usputan posao koji se obavlja kada se tekst uređuje za daljnje korištenje ili objavljivanje.

Pri redakturi poduzima se najmanji broj koraka koji je potreban da se isprave uočene pogreške, no čak i kad je neki tekst točan i tečan, on svejedno ne odgovara uvijek očekivanjima koje čitatelji imaju od prijevoda. Rečenice se u procesu prevođenja, primjerice, mogu spajati ili razlamati, što najčešće nije slučaj u strojnom prevođenju, ili bi neki prevoditelj koji uređuje strojni prijevod svejedno radije nanovo preveo neku rečenicu služeći se drugim jezičnim materijalom nego odabrao samo ispraviti pogreške strojnog prijevoda. Sve to upućuje da je analiza pogrešaka odgovarajuća metoda samo onda kada istraživač ciljano želi informaciju upravo o tome koje se vrste pogrešaka javljaju u nekom strojnom prijevodu, a ne onda kada ga zanima koliko je tekst razumljivo preveden, što je dokazano i ovim istraživanjem.

6. Zaključak

Cilj je ovog istraživanja bio usporediti različite metode vrednovanja strojnih prijevoda i proučiti u kojim se uvjetima provedbe metoda mogu dobiti najobjektivniji podaci. Za istraživanje su odabrane metode subjektivne procjene, testa razumijevanja i analize pogrešaka, a u vrednovanju je sudjelovalo sto šezdeset ispitanika i jedan profesionalni prevoditelj koji je anotirao jezične pogreške. Kako bi se podaci mogli dijakronijski pratiti, uključeni su i rezultati predistraživanja (Ljubas 2019) u kojemu je sudjelovalo još dodatnih četrdeset i petero ispitanika, a koji su vrednovali prijevod istog izvornog materijala kao i u glavnom istraživanju, samo dvije godine prije.

Ispitanici su prvo bili podijeljeni u dvije skupine; jedna je dodijelila ocjenu razumljivosti za svaki pročitani segment iz teksta, a druga je dodijelila ocjenu razumljivosti za čitav tekst. Iako se tim postupkom htjelo provjeriti u kojoj će se situaciji postići veća razina slaganja, bitnija se dosljednost u ocjenama nije zabilježila ni kod jedne skupine, ali se zato uvidjelo da je tekstove danas, kada su izlazni podaci mnogo bolji, ipak smislenije vrednovati kao cjeline, kako bi vrednovatelji mogli usmjeriti pažnju i na elemente koji nadilaze rečeničnu razinu. U drugom su zadatku ispitanici prešli na test razumijevanja čiji su se rezultati pokazali mnogo dosljednijima i objektivnijima nego što je to bio slučaj kod metode subjektivne procjene. Vrlo visok postotak točnih odgovora na određena pitanja ukazao je na to koji su dijelovi teksta osobito kvalitetno i razumljivo prevedeni. Kako te dijelove teksta ispitanici nisu nužno ocijenili boljim ocjenama, uvidjela se nasumičnost metode subjektivne procjene. Slabije ocjene nisu se mogle jednostavno otpisati ni kao posljedica negativnih stavova ispitanika jer je anketni obrazac pokazao da su ispitanici pretežno pozitivno nastrojeni prema strojnom prevođenju, a nije utvrđena ni izravan suodnos s brojem ili vrstom jezičnih pogrešaka. U analizi pogrešaka utvrđeno je pak da je potrebno uskladiti klasifikacije pogrešaka jer se u neuronskim prijevodima javlja druga vrsta pogrešaka nego što je to dosad bio slučaj.

Postavljeno je šest istraživačkih pitanja, a možda se i najrelevantnijim pokazalo upravo IP1 (*Postiže li se veće slaganje među ispitanicima kada oni ocjenjuju prijevod u cjelini u odnosu na raščlanjene segmente ili obrnuto?*) koje se uz IP2 odnosilo na metodu subjektivne analize. Kao što je već rečeno, nije se postigla bitnija razina slagana ni u jednoj od dviju skupina ispitanika, ali pobližim proučavanjem materijala i odgovora ispitanika pokazalo se da je vrednovanje raščlanjenih segmenata zastarjela metoda, koja je odgovarala pretežno statističkim

sustavima, no koja više nije prikladna kada se vrednuju rezultati neuronskih strojnih sustava. Ti su nalazi u skladu s najrecentnijim istraživanjima vrednovanja subjektivnom procjenom (v. npr. Mitchell 2019, Läubli i sur. 2020), a koji se polako odmiču od izvornih ideja o provođenju metode subjektivne procjene na kratkim, najčešće rečeničnim segmentima. Preporuka za daljnja istraživanja na neuronskim strojnim prijevodima glasi da prijevode svakako valja vrednovati u cjelini kada se rabi metoda subjektivne procjene.

Utjecaj iskustava i stavova ispitanika na metodu subjektivnog vrednovanja, koji se ispitao u okviru IP2 (*Ovise li ocjene pri procjeni razumljivosti teksta o prijašnjim iskustvima sa strojnim prevođenjem ili o stavovima o njemu?*), već je u prošlosti privukao zanimanje istraživača koji su uočili da upravo zbog utjecaja stavova ispitanika prema strojnom prevođenju i treba pronalaziti objektivnije metode vrednovanja, poput metode testova razumijevanja (v. Fuji 1999, Simeon 2008, Sanders i sur. 2011). Nalazi ovog istraživanja prvenstveno su pokazali pomak u stavovima ispitanika: dok su proteklih desetak godina često bilježeni negativni stavovi, u ovom istraživanju većina je ispitanika pokazala pozitivne stavove prema alatima za strojno prevođenje. Također se pokazalo da varijabla stavova ispitanika o strojnom prevođenju, bilo oni pozitivni ili negativni, više nije jednako ometajuća. Odgovor na IP2 odmiče se, dakle, od prethodno uočenih suodnosa – ocjene pri procjeni razumljivosti danas sve manje ovise o stavovima ispitanika. Saznanje o pretežno pozitivnim stavovima ispitanika različitih struka vrlo je važno i za hrvatske prevoditelje jer ukazuje na to da su strojni prijevodi u društvu već prisutni i prihvaćeni te da ih stoga i oni ne mogu zanemarivati i boriti se protiv tog fenomena, već ga trebaju prihvatiti i naučiti kako ga korisno primijeniti u svom profesionalnom radu, dok im s druge strane ostaje pripremiti se za borbu za bolje honorare i uvjete rada, na koje također dolazak strojnog prevođenja također utječe.

Istražujući odgovor na IP3 (*Jesu li ocjene pri procjeni razumljivosti teksta proporcionalne broju točnih odgovora na testu razumijevanja?*) nametnuo se zaključak o metodi testova razumijevanja kao najobjektivnijoj od proučenih metoda vrednovanja. Pokazalo se da se u istim uvjetima ne postižu jednaki rezultati metodom subjektivne procjene i testa razumijevanja. Premda se to konkretno pitanje ne tematizira učestalo, odnosno nije provedeno mnogo direktnih usporedbi subjektivne metode i metode testa razumijevanja kao u ovom istraživanju, ista se diskrepancija između ocjena dodijeljenih prijevodima i uspješnosti na testu pokazala i u istraživanju koje su 2018. proveli Macken i Ghyselen, pa se može reći da su nalazi u skladu s

dosadašnjom literaturom te da istraživači postaju svjesniji prednosti metode testa razumijevanja nad subjektivnom procjenom.

Odgovor IP4 (*Ovise li ocjene pri procjeni razumljivosti pojedinih segmenata teksta o broju jezičnih pogrešaka u njima?*) glasi da ne postoji direktan utjecaj broja pogreška u određenom segmentu teksta na percipiranu razumljivost tog segmenta, pa se čitavo to pitanje pokazalo manje relevantnim u kontekstu neuronskog strojnog prevođenja nego što je to bio slučaj u prošlosti. I to je vrijedno saznanje koje upućuje na to da se istraživači s neuronskim prijevodima, u kojima se bilježi sve manje morfosintaktičkih pogrešaka kakve su u statističkim modelima bile vrlo učestale pri prevođenju na hrvatski (v. npr. Ljubas 2018), susreću s nekim novim problemima, dok stari gube na važnosti. To također znači da metoda analize pogrešaka zapravo nije prilagođena kriteriju razumljivosti jer nam broj određenih vrsta pogrešaka ne govori ništa o tome koliko je pojedinom ispitaniku tekst bio razumljiv. Slična je situacija uočena i kod IP5 (*Utječu li pogreške u leksičkoj točnosti bitnije na nerazumijevanje prijevoda od ostalih pogrešaka?*), gdje također odgovor glasi da leksičke pogreške danas više nisu glavni izvor nerazumljivosti u prijevodima. To saznanje govori u korist ideje da se analize trebaju provoditi sa što manje kategorija, koje će biti prilagođene pojedinom tekstu, namjeri provođenja vrednovanja, a osobito jezičnom paru (v. Popović 2018, Klubička, Toral i Sánchez-Cartagena 2018).

Zadnje istraživačko pitanje, IP6 (*Postiže li se smanjenjem broja jezičnih pogrešaka u izlaznim podacima veća dosljednost u rezultatima vrednovanja metodama subjektivne procjene i testa razumijevanja?*), pažnju je usmjerilo na dijakronijsku usporedbu. Istraživači koji su se nakon njegova pojavljivanja bavili neuronskim strojnim prevođenjem odmah su zamijetili jasne pomake u kvaliteti prijevoda (v. npr. Bentivogli i sur. 2016, Castilho i sur. 2017, Toral i Sánchez-Cartagena 2017). Posebno je hvaljena veća lakoća čitanja prevedenih tekstova, što je posljedica smanjenog broja morfosintaktičkih, ali i leksičkih pogrešaka u neuronskim prijevodima. Ovim se istraživanjem pokazalo da i metode vrednovanja moraju slijediti taj razvoj – već se kod subjektivne metode ispostavilo da vrednovanje pruža nevjerodostojnije rezultate što su prijevodi kvalitetniji. Moguće je to protumačiti promjenom u perspektivi – kvalitetnije prijevode ispitanici i nesvjesno uspoređuju s ljudskim prijevodima i stoga svakoj pogreški posvećuju veću pažnju, kao što bi to činili i u ljudskim prijevodima, dok su u očigledno nedostatnim strojnim prijevodima pažnju usmjeravali prema onome što je sustav uopće uspio prevesti iole kvalitetno. No, ako subjektivna metoda i pada na dijakronijskom ispitu, zahtjeve

kvalitetne metode vrednovanja ispunjava objektivnija metoda testova razumijevanja. Nalazi ovog istraživanja daju jasnu prednost metodi testa razumijevanja, koja se i u dijakronijskom pogledu pokazala vjerodostojnijom, kao i da se bolje nosi s izazovima porasta kvalitete. U korist ovoj metodi govori koliko su rezultati koji se njome dobivaju konkretni i smisleni za određivanje razumljivosti, što nije uvijek slučaj s ostalim metodama.

Ovo je istraživanje osobito aktualno u svjetlu sve boljih izlaznih podataka programa za strojno prevođenje s kojima se sve metode vrednovanja ne mogu jednako dobro nositi: iz dosadašnjih se istraživanja već znalo da je to izazov za automatske metrike (kao što su još 2008. zaključili Babych i Hartley, a potvrdio i Han 2016, v. potpoglavlje 2.2.1), a rezultati ovog istraživanja pokazuju da se ista teškoća javlja i kod metode subjektivne procjene. Pritom se od strojnog prevođenja praktički iz dana u dan očekuju sve bolji rezultati i ono doista neprestano napreduje. Stoga ovi nalazi otvaraju mnoga zanimljiva istraživačka pitanja za buduća istraživanja. Bilo bi, primjerice, korisno ponoviti dio istraživanja koji se odnosi na test razumijevanja, ali s drugom skupinom ispitanika – profesionalnim prevoditeljima. Također bi se isplatilo promotriti kako se promjenom pitanja u testu razumijevanja može promijeniti i kriterij istraživanja. Premda teško da bi se testom razumijevanja mogla ciljano ispitati tečnost izlaznih podataka, točnost bi se mogla provjeriti pitanjima koja bi se specifičnije odnosila na pojedine pojmove u tekstu. Tako bi se, umjesto dubinskog razumijevanja, moglo detaljnije propitati razumijevanje unaprijed određenih koncepata iz teksta. U tu i slične svrhe mogla bi se i promijeniti ili varirati vrsta pitanja, npr. uključivši pitanja otvorenog tipa, pitanja alternativnog izbora i slično.

Na ovom je mjestu važno ponovno istaknuti da sve bolja kvaliteta ne znači da se strojno prevođenje razvija kako bi se jednoga dana u potpunosti automatizirao proces prevođenja i da će profesionalni prevoditelji postati tehnološki višak u svijetu automatskih prijevoda. Ispitivanjem stavova ispitanika procijenjeno je da su čak i laici toga svjesni: većina njih, premda su načelno pozitivno nastrojeni prema alatima za strojno prevođenje, ističe važnost profesionalnih prevoditelja za čijim bi uslugama posegnuli u svim važnim prilikama kada im je potreban prijevod. Prevoditelji se, dakle, i dalje smatraju nezamjenjivima. Laici se strojnim prijevodima služe tek u vrlo ograničenim situacijama kada trebaju brze uvide u tekstove napisane na jezicima koje sami ne govore, i u takvim su se slučajevima voljni izboriti i s njihovim nedostacima, što ne znači da ih ne zamjećuju. Zamjećuju ih svakako i prevoditelji, koji ih trebaju shvatiti kao dodatan argument u svojoj nezamjenjivosti, ali ne i kao argument za

potpuno odbacivanje svih blagodati koje uvođenje strojnog prevođenja u svakodnevni rad može donijeti. Razvoj sustava za strojno prevođenje unosi promjene u poslovne živote prevoditelja, ali na prevoditeljima je da ih pametno iskoriste kako bi si olakšali svakodnevnicu te da u poslu pronađu kreativnost i zadovoljstvo koje je nerijetko ugušeno repetitivnim i dosadnim zadacima.

Tri su glavna znanstvena doprinosa ove doktorske disertacije. Prvo, u Hrvatskoj dosad nije provedeno ovako sveobuhvatno istraživanje najrelevantnijih metoda vrednovanja strojnog prevođenja. Istraživanje je provedeno na velikom broju ispitanika i dobiveni su pouzdani rezultati o prednostima i manama metoda subjektivne procjene, testa razumijevanja i analize pogrešaka kada se provode na tekstovima prevedenima na hrvatski jezik, što istraživačima omogućuje informiraniji odabir metode za buduća istraživanja. Osobito je to važno u dobu kada kvaliteta strojnih prijevoda sve više i brže napreduje, do te mjere da zbog tečnosti u rezultatima danas ne možemo uvijek posve lako odrediti da je nešto strojni, a ne ljudski prijevod. U kontekstu istraživanja vrednovanja u Hrvatskoj također je novitet i to što je ovo istraživanje provedeno s njemačko-hrvatskim jezičnim parom, dosad manje uobičajenom jezičnom kombinacijom, premda je njemački jedan od velikih svjetskih jezika. Velika se većina vrednovanja u Hrvatskoj, kao i u svijetu, provodi u kombinaciji s engleskim jezikom, između ostalog zato što se engleski kao jezik akademije nameće kao logičan izbor i najpraktičniji je kada je u akademskim člancima ili konferencijskim izlaganjima potrebno prikazati i analizirati primjere za međunarodnu publiku.

Drugi doprinos ovog istraživanja konkretna su njegova saznanja: utvrđeno je da je metoda subjektivne procjene danas još jedino relevantna ako se vrednuju veće cjeline, a ne više raščlanjeni segmenti, da je metoda testa razumijevanja najobjektivnija ljudska metoda vrednovanja od proučenih, te da je kod metode analize pogrešaka potrebno ograničiti kategorije pogrešaka na nekoliko kategorija. Istraživanje je pratilo međunarodne trendove: utvrdila se prikladnost i važnost metrike za vrednovanje MQM koja je u trenutku provođenja istraživanja još bila relativno nova, ali se do objavljivanja ove disertacije prometnula u *de facto* standard analitičkog mjerenja kvalitete, a koji će uskoro biti proglašen i službenim¹⁸. Saznanja ovog istraživanja tako su istovremeno i relevantna i vrlo moderna, i u hrvatskom i u međunarodnom kontekstu istraživanja vrednovanja, gdje se sve češće govori o potrebnom modificiranju poznatih metoda. Uz to što je utvrđena objektivnost rezultata vrednovanja testovima

¹⁸ Kao što je autor metrike, Arle Lommel, najavio na izlaganju „Good Enough or Full of Holes? Translation Quality in a Technological World“ koje su on i Kateřina Gašová održali 10. studenog 2022. na konferenciji Translating Europe Forum 2022 održanoj u Bruxellesu te *online*.

razumijevanja, primjerom je predloženo kako što lakše pripremiti takva vrednovanja: preuzimanjem već standardiziranih testova za provjeru znanja jezika.

Nalazi ovog istraživanja zapravo se udaljavaju od starijih praksi, ali su u skladu s vrlo recentnim tendencijama i mijenama u polju vrednovanja strojnog prevođenja. Moglo bi se reći da se nalazi poklapaju s idejama koje su i u međunarodnom kontekstu još u povojima i koje istraživači strojnog prevođenja diljem svijeta tek u posljednjih godinu ili dvije formuliraju u konkretne zaključke o najboljim praksama provođenja vrednovanja strojnog prevođenja. Kao treći doprinos ovog istraživanja može se istaknuti da ono pridonosi razvijanju i etabiliranju poddiscipline vrednovanja strojnog prevođenja unutar računalne lingvistike i znanosti o prevođenju u Hrvatskoj, nadograđujući saznanja iz dosad provedenih istraživanja vrednovanja hrvatskih znanstvenika.

Na kraju, valja napomenuti da je u ovoj disertaciji proučen izbor glavnih ljudskih metoda vrednovanja kojih u stvarnosti ima mnogo više, a nikako se ne smije zaboraviti da i automatske metrike mogu biti vrlo korisne pri vrednovanju. Svaka od poznatih metoda ima svoje mjesto u povijesti razvoja vrednovanja strojnih prijevoda, pa i u praksi – metoda analize pogrešaka, primjerice, u praksi i dalje iznimno pomaže prevoditeljima u postupku redakture teksta jer zbog nje znaju na što trebaju obratiti osobitu pozornost kada dorađuju tekstove. Istraživanja ovakve vrste važna su jer se svježim uvidima u specifičnosti vrednovanja pojedinim metodama one mogu samo razviti u još bolje i kompletnije. Osobito je to važno u kontekstu istraživanja strojnog prevođenja u Hrvatskoj, u kojoj ima mjesta za još mnogo istraživanja i radova na ovu temu, kao što je otvoren prostor i za rad u praksi sa strojnim prevođenjem, koje na mala vrata ulazi u radnu svakodnevicu i ovdašnjih prevoditelja.

Literatura

1. Ageeva, Ekaterina; Forcada, Mikel L.; Tyers, Francis M. i Pérez-Ortiz, Juan Antonio. (2015). „Evaluating machine translation for assimilation via a gap-filling task“. U: *Proceedings of the 18th annual conference of the European Association for Machine Translation*. Antalya: EAMT. 137-144.
2. Aikawa, Takako i Rarrick, Spencer. (2011). „Are numbers good enough for you? A linguistically meaningful MT evaluation method“. U: *Proceedings of the 13th machine translation summit (AAMT-2011)*. Xiamen: CIPS i XMU. 332-337.
3. Allen, Jeffrey. (2003). „Post-editing“. U: Somers, Harold (ur.) *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins. 297-318.
4. ALPAC. (1966). *Languages and machines: computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee Division of Behavioral Sciences National Academy of Sciences National Research Council*. Washington D. C.: National Academy of Sciences, National Research Council.
5. Arnold, Douglas. (2003). „Why translation is difficult for computers“. U: Somers, Harold (ur.) *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins. 119-142.
6. Arnold, Douglas; Balkan, Lorna; Meijer, Siety; Humphreys, R. Lee i Sadler, Louisa. (1994). *Machine Translation: an introductory guide*. London: Blackwell.
7. Babych, Bogdan i Hartley, Anthony. (2008). „Sensitivity of automated MT evaluation metrics on higher quality MT output: BLEU vs task-based evaluation methods“. U: Calzolari, Nicoletta; Choukri, Khalid; Maegaard, Bente; Mariani, Joseph; Odijk, Jan; Piperidis, Stelios i Tapias, Daniel (ur.) *Proceedings of the 6th international conference on language resources and evaluation (LREC'08)*. Marrakech: ELRA. 2133-2136.

8. Baker, Jim; Brown, Peter; Carlson, Lynn; Hovy, Eduard; Wayne, Charles i White, John. (1992). *Machine Translation Evaluation Methodology*. URL: www.mt-archive.info/DARPA-1992-Baker.pdf. Pristup: 11. srpnja 2018.
9. Bayerl, Petra Saskia i Paul, Karsten Ingmar (2011). „What determines inter-coder agreement in manual annotations? A meta-analytic investigation“. U: Merlo, Paola (ur.) *Computational Linguistics* 37(4): 699-725.
10. Benkova, Lucia; Munková, Daša; Benko, Lubomir i Munk, Michal. (2021). „Evaluation of English-Slovak Neural and Statistical Machine Translation“. U: Szymanski, Julian (ur.) *Applied Sciences* 11: 1-17.
11. Berka, Jan; Černý; Martin i Bojar, Ondřej. (2011). „Quiz-based evaluation of machine translation“. U: Hajičová, Eva (ur.) *The Prague bulletin of mathematical linguistics* 95(1). 77-86.
12. Bennett, Winfield Scott i Gerber, Laurie. (2003). „Inside commercial machine translation“. U: Somers, Harold (ur.) *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins. 175-190.
13. Bentivogli, Luisa; Bisazza, Arianna; Cettolo, Mauro i Federico, Marcello. (2016). „Neural versus phrase-based machine translation quality: A case study“. U: Su, Jian; Duh, Kevin i Carreras, Xavier (ur.) *Proceedings of the 2016 conference on empirical methods in natural language processing*. Austin: ACL. 257-267.
14. Bojar, Ondřej. (2011). „Analyzing error types in English-Czech machine translation“. U: Hajičová, Eva (ur.) *The Prague bulletin of mathematical linguistics* 95(1). 63-76.
15. Bowker, Lynne i Buitrago-Ciro, Jairo. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Bingley: Emerald publishing.

16. Burchardt, Aljoscha i Lommel, Arle (2014). *Practical guidelines for the use of MQM in scientific research on translation quality*. Dostupno na: <http://www.qt21.eu/downloads/MQM-usage-guidelines.pdf>. Pristup: 8. svibnja 2020.
17. Bremin, Sofia; Hu, Hongzhan; Karlsson, Johanna; Prytz Lillkull, Anna; Wester, Martin; Danielsson, Henrik i Stymne, Sara (2018). „Methods for human evaluation of machine translation“. U: *Proceedings of the Swedish language technology conference (SLTC2010)*. Linköping: Linköpings universitet. 47-48.
18. Callison-Burch, Chris; Fordyce, Cameron Shaw; Koehn, Philipp; Monz, Christof i Schroeder, Josh. (2008). „Further meta-evaluation of machine translation“. U: Callison-Burch, Chris; Koehn, Philipp; Monz, Christof; Schroeder, Josh i Fordyce, Cameron Shaw. (ur.) *Proceedings of the 3rd workshop on statistical machine translation (WMT-08)*. Columbus: ACL. 70-106.
19. Carroll, John B. (1966). „An experiment in evaluating the quality of translations“. U: *Mechanical translation and computational linguistics* 9(3/4). 55-66.
20. Castilho, Sheila; Moorkens, Joss; Gaspari, Federico; Calixto, Iacer; Tinsley, John i Way, Andy. (2017). „Is neural machine translation the new state of the art?“. U: Hajič, Jan (ur.) *The Prague bulletin of mathematical linguistics* 108(1): 109-120.
21. Castilho, Sheila i Guerberof Arenas, Ana. (2018). „Reading comprehension of machine translation output: What makes for a better read?“. U: Pérez-Ortiz, Juan Antonio; Sánchez-Martínez, Felipe; Esplá-Gomis, Miquel; Popović, Maja; Rico, Celia; Martins, André; Van den Bogaert, Joachim i Forcarda, Mikel L. (ur.) *Proceedings of the 21st annual conference of the European Association for Machine Translation*. Alacant: EAMT. 79-88.

22. Castilho, Sheila; Doherty, Stephen; Gaspari, Federico i Moorkens, Joss. (2018). „Approaches to human and machine translation quality assessment“. U: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico i Doherty, Stephen. (ur.) *Translation quality assessment: From principles to practice*. Cham: Springer. 9-38.
23. Day, Richard R. i Park, Jeongyeon. (2005). „Developing reading comprehension questions“. U: Day, Richard R. i Macalister, John (ur.) *Reading in a foreign language*, 17(1). 60-73.
24. Doherty, Stephen. (2017). „Issues in human and automatic translation quality assessment“. U: Kenny, Dorothy (ur.) *Human issues in translation technology*. London: Routledge. 131-148.
25. Dorr, Bonnie; Snover, Matthew i Madnani, Nitin. (2011). „Machine translation evaluation and optimization: Introduction“. U: Olive, Joseph; Christianson, Caitlin i McCary, John (ur.) *Handbook of natural language processing and machine translation*. New York: Springer. 745-747.
26. Dovedan, Zdravko; Seljan, Sanja i Vučković, Kristina. (2002). „Strojno prevođenje kao pomoć u procesu komunikacije“. U: Plenković, Mario (ur.) *Informatologia* 35(4). 283-291.
27. Drugan, Joanna. (2013). *Quality in professional translation: assessment and improvement*. London: Bloomsbury.
28. Fan, Angela; Bhosale, Shruti; Schwenk, Holger; Ma, Zhiyi; El-Kishky, Ahmed; Goyal, Siddharth; Baines, Mandeep; Celebi, Onur; Wenzek, Guillaume; Chaudhary, Vishrav; Goyal, Naman; Birch, Tom; Liptchinsky, Vitaly; Edunov, Sergey; Grave, Edourad; Auliy, Michael i Jouli, Armand (2021.) „Beyond English-centric multilingual machine translation“. U: Clark, Alexander (ur.) *Journal of machine learning research* 22(2021). 1-48.

29. Federico, Marcello; Negri, Matteo; Bentivogli, Luisa i Turchi, Marco. (2014). „Assessing the impact of translation errors on machine translation quality with mixed-effects models“. U: Moschitti, Alessandro; Pang, Bo i Daelemans, Walter (ur.) *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP 2014)*. Doha: ACL. 1643-1653.
30. Freitag, Markus; Foster, George; Grangier, David; Ratnakar, Viresh; Tan, Qijun i Macherey, Wolfgang. (2021). „Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation“. U: Nenkova, Ani i Roark, Brian (ur.) *Transactions of the Association for Computational Linguistics* 9(1). 1460-1474
31. Fuji, Masaru. (1999). „Evaluation experiment for reading comprehension of machine translation outputs“. U: *Proceedings of the 7th machine translation summit: MT in the great translation era*. Singapur. 285-289.
32. Fuji, Masaru; Hatanaka, Nobutoshi; Ito, Etsuo; Kamei, Shin-ichiro; Kumai, Hiroyuki; Sukehiro, Tatsuya; Yoshimi, Takehiko i Isahara, Hitoshi. (2001). „Evaluation method for determining groups of users who find MT 'useful'“. U: Maegaard, Bente (ur.) *Proceedings of the 8th machine translation summit: Machine translation in the information age*. Santiago de Compostela: EAMT. 103-108.
33. Ghyselen, Iris. (2017). *Assesing the comprehensibility and perception of machine translation*. Doktorska disertacija. Gent: Universiteit Gent.
34. Girardi, Christian; Bentivogli, Luisa; Farajian, Mohammad Amin; Federico, Marcello. (2014). „MT-EquAl: A toolkit for human assessment of machine translation output“. U: Tounsi, Lamia i Rak, Rafal (ur.) *Procedings of COLING 2014: the 25th international conference on computational linguistics: System Demonstrations*. Dublin: Dublin City University i ACL. 120-123.

35. Göpferich, Susanne. (2009). „Comprehensibility assessment using the Karlsruhe comprehensibility concept“. U: Byrne, Jody (ur.) *The journal of specialised translation*, 11(2009). 31-51.
36. Gouadec, Daniel. (2010). „Quality in translation“. U: Gambier, Yves i van Doorslaer, Luc (ur.) *Handbook of translation studies*. Amsterdam: Benjamins. 270-275.
37. Guerberof Arenas, Ana i Toral, Antonio. (2022). „Creativity in translation: Machine translation as a constraint for literary texts“. U: Kenny, Dorothy i Moorkens, Joss (ur.): *Translation Spaces, Online First-articles*. Amsterdam: John Benjamins.
38. Gupta, Vaishali; Joshi, Nisheeth i Mathur, Iti. (2013). „Subjective and objective evaluation of English to Urdu machine translation“. U: *Proceedings of the 2013 international conference on advances in computing, communications and informatics (ICACCI)*. Mysore: SJCE. 1520-1525.
39. Guzmán, Francisco; Abdelali, Ahmed; Temnikova, Irina; Sajjad, Hassan i Vogel, Stephan. (2015). „How do humans evaluate machine translation“. U: Bojar, Ondřej; Chatterjee, Rajan; Federmann, Christian; Haddow, Barry; Hokamp, Chris; Huck, Matthias; Logacheva, Varvara i Pecina, Pavel. (ur.) *Proceedings of the 10th workshop on statistical machine translation (WMT-15)*. Lisabon: ACL. 457-466.
40. Han, Lifeng. (2016). *Machine translation evaluation resources and methods: A survey*. Dostupno na: <https://arxiv.org/abs/1605.04515>. Pristup: 7. svibnja 2020.
41. Hansen-Kokoruš, Renate; Matešić, Josip; Pečur-Medinger, Zrinka i Znika, Marija. (2005). *Deutsch-Kroatisches Universalwörterbuch / Njemačko-hrvatski univerzalni rječnik*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.
42. Hutchins, John. (2012). „Recent applications of machine translation“. U: Malmkjær, Kirsten i Windle, Kevin (ur.) *The Oxford handbook of translation studies*. Oxford. 293-301.

43. Hutchins, John i Somers, Harold. (1992). *An introduction to machine translation*. London: Academic Press.
44. ISO. (2017). *ISO 18857:2017: Translation services – post-editing of machine translation output: Requierments*.
Dostupno na: <https://www.iso.org/standard/62970.html>. Pristup: 1. prosinca 2023.
45. Jones, Douglas; Herzog, Martha; Ibrahim, Hussny; Jairam, Arvind; Shen, Wade; Gibson, Edward i Emonts, Michael. (2007). „ILR-based MT comprehension test with multi-level questions“. U: *Proceedings of NAACL HLT 2007 – Companion volume*. Rochester: ACL. 77-80.
46. Kenny, Dorothy. (2022). „Human and machine translation“. U: Kenny, Dorothy (ur.) *Machine translation for everyone – Empowering users in the age of artificial intelligence*. 23-50.
47. King, Margaret; Popescu-Belis, Andrei i Hovy, Eduard. (2003). „FEMTI: Creating and using a framework for MT evaluation“. U: *Proceedings of the 9th machine translation summit*. New Orleans. 224-232.
48. Klubička, Filip; Toral, Antonio i Sánchez-Cartagena, Víctor M. (2018). „Quantitative fine-grained human evaluation of machine translation systems: a case study on English to Croatian“. U: Way, Andy i Forcada, Mikel L. (ur.) *Machine Translation* 32(3). 195-215.
49. Koehn, Philipp. (2009). *Statistical machine translation*. Cambridge: University Press.
50. Koeh, Philipp. (2020). *Neural machine translation*. Cambridge: University Press.
51. Kolb, Waltraud. (2022). „Welche Rolle können Maschinen in der Literaturübersetzung spielen?“. U: Palundo, Tamara (ur.) *Universitas-Mitteilungsblatt* 1/22. 19-23.

52. Koletnik Korošec, Melita. (2011). „The Internet, Google Translate and Google Translator Toolkit – Nuisance or necessity in translator training?“. U: *Proceedings of Tralogy I: Futures in technologies for translation*. Paris. Dostupno na: <http://lodel.irevues.inist.fr/tralogy/index.php?id=113>. Pristup: 13. svibnja 2020.
53. Koponen, Maarit. (2012). „Comparing human perceptions of post-editing effort with post-editing operations“. U: Callison-Burch, Chris; Koehn, Philipp; Monz, Christof; Post, Matt; Soricut, Radu i Specia, Lucia (ur.) *Proceedings of the 7th workshop on statistical machine translation*. Montréal: ACL. 181-190.
54. Kresić, Marijana i Batinić, Mia. (2014). *Modalpartikeln: Deutsch im Vergleich mit dem Kroatischen und Englischen. Modalne čestice: njemački jezik u usporedbi s hrvatskim i engleskim*. Zadar: Sveučilište u Zadru.
55. Kresić Vukosav, Marijana i Batinić Angster, Mia. (2019). *Višejezični priručnik za učenje modalnih čestica: njemački, hrvatski, engleski. Mehrsprachiger Leitfaden zum Erlernen von Modalpartikeln: Deutsch, Kroatisch, Englisch*. Zadar: Morepress – Sveučilište u Zadru.
Dostupno na: <https://morepress.unizd.hr/books/index.php/press/catalog/book/41>.
Pristup: 29. rujna 2022.
56. Krings, Hans P. (2001). *Repairing texts: Empirical investigations of machine translation post-editing process*. Kent: The Kent State University Press.
57. Kučiš, Vlasta i Seljan, Sanja (2014). „The role of online translation tools in language education“. U: De Laet, Frans (ur.) *Babel 60(3)*. 303-324.
58. Kučiš, Vlasta; Seljan, Sanja i Klasnić, Ksenija. (2009). „Evaluation of electronic translation tools through quality parameters“. U: Stančić, Hrvoje; Seljan, Sanja; Bawden, David; Lasić-Lazić, Jadranka i Slavić, Aida (ur.) *INFuture2009: Digital resources and knowledge sharing – Proceedings*. Zagreb: Odsjek za informacijske znanosti, Filozofski fakultet, Sveučilište u Zagrebu. 341-351.

59. Läubli, Samuel; Castilho, Sheila; Neubig, Graham; Sennrich, Rico; Shen, Qinlan i Toral, Antonio. (2020.) „A set of recommendations for assessing human-machine parity in language translation“. U: Markovitch, Shaul (ur.) *Journal of artificial intelligence research* 67(2020). 653-672.
60. Lommel, Arle; Gašová, Kateřina. (2022, 9. – 11. studenog). *Good Enough or Full of Holes? Translation Quality in a Technological World* [konferencijsko izlaganje]. Europe Translate Forum 2022. Bruxelles. Dostupno na:
https://www.youtube.com/watch?v=DRO1c7rLAyA&list=PLLqIRaiVCGCSrlTMx1iG1pRlyYR1ZKSyi&index=22&ab_channel=TranslatingforEurope.
Pristup: 2. prosinca 2022.
61. Lommel, Arle; Popović, Maja i Burchardt, Aljoscha. (2014). „Assessing inter-annotator agreement for translation error annotation“. U: *Proceedings of MTE Workshop on automatic and manual metrics for operational translation evaluation (LREC 2014)*. Reykjavik: ELRA. 31-37.
62. Ljubas, Sandra. (2017). „Analiza pogrešaka u strojnim prijevodima sa švedskog na hrvatski“. U: Pavlović, Nataša (ur.) *Hieronymus – časopis za istraživanja prevođenja i terminologije* 4. 28-64.
63. Ljubas, Sandra. (2018). „Prijelaz sa statističkog na neuronski model: Usporedba izlaznih podataka u prijevodima sa švedskog na hrvatski jezik“. U: Pavlović, Nataša (ur.) *Hieronymus – časopis za istraživanja prevođenja i terminologije* 5. 72-91.
64. Ljubas, Sandra. (2019). Neobjavljeno predistraživanje.
65. Ljubas, Sandra. (2020). „Utjecaj višejezičnosti vrednovatelja na ljudsku procjenu kvalitete strojnih prijevoda“. U: Buljan, Gabrijela (ur.) *Jezikoslovlje* 21(2). 207-235.

66. Macken, Lieve i Ghyselen, Iris. (2018). „Measuring comprehension and perception of neural machine translated texts: a pilot study“. U: Chambers, David; Drugan, Joanna; Esteves-Ferreira, João; Macan, Juliet Margaret; Mitkov, Ruslan i Stefanov, Olaf-Michael (ur.) *Proceedings of the 40th edition of the annual Translating and the computer conference*. 120-126.
67. Massardo, Isabella; Van der Meer, Jaap; O'Brien, Sharon; Hollowood, Fred; Aranberri, Norra i Drescher, Katrin. (2016). *MT Post-editing Guidelines*. Amsterdam: TAUS Signature Editions.
68. Maney, Tucker; Sibert, Linda; Perzanowski, Dennis; Gupta, Kalyan i Schmidt-Nielsen, Astrid. (2012). „Toward determining the comprehensibility of machine translations“. U: Williams, Sandra, Siddharthan, Advait i Nenkova, Ani (ur.) *Proceedings of the 1st workshop on predicting and improving text readability for target reader populations*. Montréal: ACL. 1-7.
69. Meurers, Detmar; Ziai, Ramon; Ott, Niels i Kopp, Janina. (2011). „Evaluating answers to reading comprehension questions in context: Results for German and the role of information structure“. U: Padó, Sebastian i Thater, Stefan (ur.) *Proceedings of the TextInfer 2011 Workshop on textual entailment*. Edinburgh: ACL. 1-9.
70. Mitchell, Melanie. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus and Giroux.
71. Nießen, Sonja; Och, Franz Josef; Leusch, Gregor i Ney, Hermann. (2000). „An evaluation tool for machine translation: Fast evaluation for MT research“. U: Gavriliidou, Maria; Carayannis, George; Markantonatou, Stella; Piperidis, Stelios i Stainhauer, Gregory (ur.) *Proceedings of the second international conference on language resources and evaluation (LREC'00)*. Atena: ELRA. Dostupno na: <http://www.lrec-conf.org/proceedings/lrec2000/pdf/278.pdf>. Pristup: 7. svibnja 2020.

72. Nomura, Hirosato. (1993). „Evaluation method of machine translation: From the viewpoint of natural language processing“. U: *Proceedings of the 4th machine translation summit: International cooperation for global communication*. Kobe. 205-207.
73. Nyberg, Eric; Mitamura, Teruko i Huijsen, Willem-Olaf. (2003). „Controlled Language for Authoring and Translation“. U: Somers, Harold (ur.) *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins. 245-282.
74. O'Brien, Sharon. (2022a). „How to deal with errors in machine translation: Post-editing“. U: Kenny, Dorothy (ur.) *Machine translation for everyone – Empowering users in the age of artificial intelligence*. 105-120.
75. O'Brien, Sharon. (2022b). „Authenticity and machine translation“. U: Mavrič, Valter (ur.) *The many faces of translation – Machine translation: driven by humans, powered by technology*. Europski parlament, Glavna uprava za pismeno prevođenje. 63-77.
76. Och, Franz. (2012). *Breaking down the language barrier – six years in*. Dostupno na: <https://googleblog.blogspot.com/2012/04/breaking-down-language-barriersix-years.html> Pristup: 8. svibnja 2020.
77. Okpor, Margaret Dumebi. (2014). „Machine translation approaches: Issues and challenges“. U: Vanrullen, Tristan (ur.) *International journal of computer science issues (IJCSI)*, 11(5). 159-165.
78. Pavlović, Nataša. (2015). *Uvod u teorije prevođenja*. Zagreb: Leykam international.
79. Pavlović, Nataša. (2017). „Strojno i konvencionalno prevođenje s engleskoga na hrvatski: usporedba pogrešaka“. U: Stolac, Diana i Vlastelić, Anastazija (ur.) *Jezik kao predmet proučavanja i jezik kao predmet poučavanja*. Zagreb: Srednja Europa. 279-295.

80. Pérez-Ortiz, Antonio; Forcada, Mikel L. i Sánchez-Martínez, Felipe. (2022). „How neural machine translation works“. U: Kenny, Dorothy (ur.) *Machine translation for everyone – Empowering users in the age of artificial intelligence*. 141-164.
81. Popović, Maja. (2018). „Error classification and analysis for machine translation quality assessment“. U: Moorkens, Joss; Castilho, Sheila; Gaspari, Federico i Doherty, Stephen (ur.) *Translation quality assessment: From principles to practice*. Cham: Springer. 129-158.
82. Pym, Peter. (1990). „Pre-editing and the use of simplified writing for MT: An engineer’s experience of operating an MT system“. U: Mayorcas, Pamela (ur.) *Translating and the computer 10: The translation environment 10 years on*. London: Aslib. 80-95.
83. Rivera-Trigueros, Irene. (2022). „Machine translation systems and quality assessment: a systematic review“. U: Ide, Nancy i Calzolari, Nicoletta (ur.) *Language Resources and Evaluation, 2022(56)*. 593-619.
84. Rossi, Caroline i Carré, Alice. (2022). „How to choose a suitable neural machine translation solution“. U: Kenny, Dorothy (ur.) *Machine translation for everyone – Empowering users in the age of artificial intelligence*. 51-80.
85. Sachs, Jacqueline. (1967). „Recognition memory for syntactic and semantic aspects of connected discourse“. U: *Perception & Psychophysics, 1967(2)*. 437-442.
86. Sanders, Gregory; Przybocki, Mark; Madnani, Nitin i Snover, Matthew. (2011). „Machine translation evaluation and optimization: Human subjective judgments“. U: Olive, Joseph; Christianson, Caitlin i McCary, John (ur.) *Handbook of natural language processing and machine translation..* New York: Springer. 750-758.

87. Savoldi, Beatrice; Gaido, Marco; Bentivogli, Luisa; Negri, Matteo i Turchi, Marco. (2021). „Gender Bias in Machine Translation“. U: Nenkova, Ani i Roark, Brian (ur.) *Transactions of the Association for Computational Linguistics 2021*. 845-874.
Dostupno na:
https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00401/106991/Gender-Bias-in-Machine-Translation. Pristup: 28. listopada 2021.
88. Scarton, Carolina i Specia, Lucia. (2016). „A reading comprehension corpus for machine translation evaluation“. U: Calzolari, Nicoletta; Choukri, Khalid; Declerck, Thierry; Goggi, Sara; Grobelnik, Marko; Maegaard, Bente; Mariani, Joseph; Mazo, H  l  ne; Moreno, Asunci  n; Odi  k, Jan i Piperidis, Stelios (ur.) *Proceedings of the 10th international conference on language resources and evaluation (LREC'16)*. Portoro  : ELRA. 3652-3658.
89. Sch  fer, Falko. (2003). „MT post-editing: How to shed light on the ‘unknown task’. Experiences made at SAP“. U: *Proceedings of the EAMT-CLAW '03: joint conference combining the 8th International workshop of the European Association for Machine Translation and the 4th Controlled language applications workshop: controlled language translation*. Dublin: EAMT. 133-140.
90. Seljan, Sanja; Klasni  c, Ksenija; Stojanac, Mara; Pe  orda, Barbara i Mikeli  c Preradovi  c, Nives. (2015). „Information transfer through online summarizing and translation technology“. U: Anderson, Karen; Duranti, Luciana; Jaworski, Rafa  ; Mickiewicz, Adam; Stan  i  c, Hrvoje; Seljan, Sanja i Mateljan, Vladimir. (ur.) *INFuture2015:e-Institutions – Openness, accessibility, and preservation – Proceedings*. Zagreb: Odsjek za informacijske znanosti, Filozofski fakultet, Sveu  ili  te u Zagrebu. 197-210.
91. Seljan, Sanja; Tucakovi  c, Marko i Dunder, Ivan. (2015). „Human evaluation of online machine translation services for English/Russian-Croatian“. U: Rocha,   lvaro; Correia, Ana Maria; Costanzo, Sandra i Reis, Luis Paulo (ur.) *Advances in Intelligent Systems and Computing 353 – New Contributions in Information Systems and Technologies*. Cham; Springer. 1089-1098.

92. Simeon, Ivana. (2008). Vrednovanje strojnoga prevođenja. Doktorska disertacija. Zagreb: Sveučilište u Zagrebu.
93. Somers, Harold. (2003). „Translation technologies and minority languages“. U: Somers, Harold (ur.) *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins. 87-104.
94. Somers, Harold. (2012). „Machine translation: History, development and limitations“. U: Malmkjær, Kirsten i Windle, Kevin (ur.) *The Oxford Handbook of Translation Studies*. Oxford. 284-292.
95. Somers, Harold i Wild, Elizabeth. (2000). „Evaluating machine translation: The cloze procedure revisited“. U: *Proceedings of the 22nd international conference on translating and the computer*. Dostupno na: <http://mt-archive.info/Aslib-2000-Somers.pdf>. Pristup: 7. svibnja 2020.
96. Stymne, Sara i Ahrenberg, Lars. (2012). „On the practice of error analysis for machine translation evaluation“. U: Calzolari, Nicoletta; Choukri, Khalid; Declerck, Thierry; Doğan, Mehmet Uğur; Maegaard, Bente; Mariani, Joseph; Moreno, Asunción; Odijk, Jan; i Piperidis, Stelios (ur.) *Proceedings of the 8th international conference on language resources and evaluation (LREC'12)*. Istanbul: ELRA. 1785-1790.
97. Sudoh, Katsuhito; Takahashi, Kosuke i Nakamura, Satoshi. (2021). „Is this translation error critical?: Classification-based human and automatic machine translation evaluation focusing on critical errors“. U: Belz, Anya; Agarwal, Shubham; Graham, Yvette; Reiter, Ehud; Shimorina, Anastasia (ur.) *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval)*. ACL. 46-55.
98. Šonje, Jure (ur.) (2000). Rječnik hrvatskog jezika. Zagreb: Leksikografski zavod Miroslav Krleža i Školska knjiga.

99. Tan, Zhixing; Wang, Shuo; Yang, Zonghan; Chen, Gang; Huang, Xuancheng; Sun, Maosong i Liu, Yang. (2020). „Neural machine translation: A review of methods, resources, and tools“. U: Tang, Jie (ur.) *AI Open* 2020(1). 5-21.
100. Toral, Antonio i Sánchez-Cartagena, Víctor M. (2017). „A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions“. U: Lapata, Mirella; Blunsom, Phil i Koller, Alexander (ur.) *Proceedings of the 15th conference of the European chapter of the association for computational linguistics: Volume 1*. 1063-1073.
101. Tripathi, Sneha i Sarkhel, Juran Krishna (2011). „Approaches to machine translation“. U: Mahesh, G. (ur.) *Annals of library and information studies* (57). 388-393.
102. Vasconcellos, Muriel. (1993). „Evaluation method of machine translation“. U: *Proceedings of the 4th machine translation summit: International cooperation for global communication*. Kobe. 211-214.
103. Yusof, Norwati; Darus, Saadiyah i Aziz, Mohd Juzaidin. (2017). „Evaluating intelligibility in human translation and machine translation“. U: *3L: The southeast Asian journal of English language studies* 23(4). 251-264.
104. Wagner, Elizabeth. (1985). „Post-editing SYSTRAN, a challenge for commission translators“. U: *Terminologie et Traduction 3*. Luxembourg: Office des publications officielles des Communautés européennes. 1-7.
105. Way, Andy. (2018). „Quality expectations of machine translation“. U: Moorkens, Joos; Castilho, Sheila; Gaspari, Federico i Doherty, Stephen (ur.) *Translation quality assessment: From principles to practice*. Cham: Springer. 159-178.

106. Way, Andy. (2020). „Machine translation: Where are we at today?“. U: Angelone, Erik; Massey, Gary i Ehrensberger-Dow, Maureen (ur.) *The Bloomsbury companion to language industry studies*. London: Bloomsbury. 311-332.
107. White, John S. (2003). „How to evaluate machine translation“. U: Somers, Harold (ur.) *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins. 211-244.
108. Wu, Yonghui; Schuster, Mike; Chen, Zhifeng; Le, Quoc V.; Norouzi, Mohammad i sur. (2016). „Google's neural machine translation system: Bridging the gap between human and machine translation.“ Dostupno na: <https://arxiv.org/pdf/1609.08144v2.pdf>. Pristup: 8. svibnja 2020.

Mrežni izvori:

109. https://bfu.goethe.de/c2_mod/lesen.php. Pristup: 6. lipnja 2021.

Sažetak

U ovome se radu analiziraju i uspoređuju tri ljudske metode vrednovanja strojnih prijevoda: metoda subjektivne procjene, metoda testa razumijevanja i metoda analize pogrešaka. Kao kriterij vrednovanja strojnih prijevoda odabrana je *razumljivost* jer su u istraživanju, umjesto prevoditelja ili lingvista, za ispitanike odabrani prosječni korisnici javno dostupnih sustava za strojno prevođenje, koji bi se njima potencijalno mogli poslužiti da shvate glavnu bit nekog teksta napisanog na jeziku koji ne razumiju, odnosno u svrhu asimilacije podataka. Njemački je izvornik na hrvatski preveden alatom Google Prevoditelj. U istraživanju je sudjelovalo sto šezdeset jednojezičnih ispitanika, koji ne govore njemački jezik i nemaju prethodnog iskustva s redigiranjem strojnih prijevoda. Ishodišna istraživačka pitanja odnosila su se na utjecaj različitih varijabli na provedbu odabranih metoda: Postiže li se veće slaganje među ispitanicima kada ocjenjuju prijevod u cjelini u odnosu na raščlanjene segmente? Ovisi li ocjene pri procjeni razumljivosti o prethodnim iskustvima sa strojnim prevođenjem ili o stavovima o njemu? Jesu li te ocjene razmjerne broju točnih odgovora na testu razumijevanja? Ovisi li o broju jezičnih pogrešaka u segmentima? Utječu li pogreške u leksičkoj točnosti značajnije na nerazumijevanje od ostalih pogrešaka? Postiže li se smanjenjem broja jezičnih pogrešaka veća dosljednost u rezultatima vrednovanja različitim metodama?

Ispitanici su u istraživanju pristupili raznovrsnim zadacima kako bi se usporedio proces primjene triju odabranih metoda vrednovanja te se utvrdilo kako se uz promjene postavki istraživanja mogu dobiti što relevantniji i objektivniji podaci o kvaliteti strojnih prijevoda. U prvom su zadatku ispitanici podijeljeni u dvije skupine: osamdeset ispitanika subjektivno je na ljestvici od 1 do 5 procijenilo razumljivost raščlanjenih i nasumično raspoređenih segmenata strojnog prijevoda, a preostalih je osamdeset ispitanika prema istoj ljestvici vrednovalo razumljivost strojnog prijevoda u cjelini. U drugom zadatku svih je sto šezdeset ispitanika čitalo cjelovit strojni prijevod te je potom odgovaralo na pitanja razumijevanja, kao i procijenilo svoju sigurnost u ponuđene odgovore. Ispitanici su mjerili vrijeme potrebno za rješavanje testa. Zatim su pristupili anketnom obrascu o stavovima i iskustvima sa strojnim prevođenjem. Analizom stavova provjeravao se njihov suodnos s ocjenama razumljivosti i uspješnosti na testu razumijevanja, te se procjenjivala objektivnost rezultata pojedinih metoda. U završnoj fazi istraživanja angažiran je stručni anotator koji je proveo analizu pogrešaka prema sljedećim kategorijama: izostavljene riječi, leksičke pogreške, morfosintaktičke pogreške, ortografske

pogreške, pogrešan redoslijed riječi i ostalo. Također su dijakronijski uspoređeni pomaci u kvaliteti izlaznih podataka u ovom istraživanju u odnosu na predistraživanje provedeno 2019. godine.

Dobiveni rezultati pokazali su da metodu subjektivne procjene valja provoditi vrednujući cjelokupni tekst radije nego raščlanjene segmente teksta, da ispitanici imaju sve pozitivnija iskustva sa strojnim prijevodima te da stavovi ispitanika danas rjeđe negativno interferiraju u procjene razumljivosti, kao i da se s metodom testa razumijevanja dobivaju objektivniji rezultati vrednovanja u odnosu na druge proučene metode. Kad je riječ o metodi analize pogrešaka, nije utvrđena direktan suodnos između broja pogrešaka u određenim segmentima i ocjena razumljivosti, a također se zaključuje da leksičke pogreške u neuronskim strojnim prijevodima nisu glavni izvor nerazumljivosti, kao što je često bio slučaj sa statističkim sustavima. Dijakronijska je analiza pokazala da su izlazni podaci danas sve veće jezične kvalitete, a da se s izazovom porasta kvalitete najbolje nosi metoda testa razumijevanja. Ovi su nalazi na tragu i najrecentnijih međunarodnih spoznaja o vrednovanju strojnog prevođenja u svijetu te predstavljaju važne doprinose istraživanju strojnog prevođenja u Hrvatskoj.

Ključne riječi: strojno prevođenje; metode vrednovanja; kriterij razumljivosti; subjektivna procjena; test razumijevanja; analiza pogrešaka

Abstract

The objective of this study was to compare three human methods for evaluating machine translations: subjective judgement, reading comprehension test and error analysis. The machine translations were evaluated against the intelligibility criterion. The respondents selected for our study were neither translators nor linguists, but average users of publicly accessible machine translation systems who usually want to grasp the gist of a text in another language, i.e. who use it for assimilation. The source text was translated from German into Croatian using Google Translate. A total of 160 monolingual respondents without any prior experience in the post-editing of machine translations participated in the study. In this study, we focused on various variables that can impact MT-evaluation. The research questions raised were: Can a higher inter-annotator agreement be attained if the respondents evaluate isolated segments or whole texts? Do subjective judgements depend on previous experiences with MT and/or the evaluators' attitudes towards MT? Are their intelligibility grades proportionate to the respondents' scores on the reading comprehension test? Are they influenced by the number of language errors in particular segments? Do lexical errors have a more significant impact on unintelligibility than other error types? Does evaluating segments with fewer errors automatically lead to a higher inter-annotator agreement?

The respondents were given a series of tasks, with the idea of finding the most relevant and objective evaluation method. In the first phase, the respondents were divided into two groups: a total of 80 respondents evaluated the intelligibility of isolated and randomly presented segments on a scale from 1 to 5, and the remaining 80 respondents evaluated the intelligibility of the whole machine translation. In the second phase, as many as 160 evaluators read the entire target text, proceeded to take a reading comprehension test, assessed how confident they were in their answers, as well as measured the time needed to complete the test. The respondents then filled out a perception survey on their attitudes and prior experience with MT. These data were compared with the subjective intelligibility judgements and the respondents' performance on the reading comprehension test in order to determine the objectivity of the individual methods. In the last phase, an expert annotator was asked to conduct an error analysis with regard to the following categories: omitted words, lexical errors, morphological errors, orthographic errors, word-order errors, and "miscellaneous". Lastly, changes in the output data quality were diachronically compared with the results of our pilot study conducted in 2019.

Several conclusions can be drawn from this study. First, it is better to subjectively evaluate the entire translation instead of individual segments. Second, the respondents generally have a rather positive attitude towards MT and their attitudes do not interfere with the subjective evaluation as significantly as in the previous years. Furthermore, based on our findings and discussion, reading comprehension tests lead to more objective results in comparison with other evaluation methods. As for the error analysis, no direct parallel could be drawn between the intelligibility scores and the number of errors in segments. There is also no indication that lexical errors would be the primary source of unintelligibility in neural machine translation, which was usually the case with statistical machine translation. The diachronic analysis showed an increase in the quality of MT-output data, but even though evaluating MT-output of high quality is challenging for many metrics, the reading comprehension test method seems to handle it most adequately out of the three studied methods. Our findings are consistent with the findings of recent international studies on evaluating MT and serve as a strong contribution to the field of MT-evaluation studies in Croatia.

Keywords: machine translation; evaluation methods; intelligibility; subjective judgements; reading comprehension tests; error analysis

	mir keinen besseren Beruf vorstellen.	mogao zamisliti bolji posao.		
12	Eigentlich müsste ich mich mit meinem Beruf nicht verstecken, könnte stolz auf ihn sein.	Zapravo, ne bih morao skrivati svoj posao, mogao bih biti ponosan na njega.	Zapravo, ne bih se trebao skrivati sa svojim poslom, mogao bih biti ponosan na njega.	Zapravo uopće ne bih trebala skrivati da sam nastavnica, trebala bih moći biti ponosna na to.
13	Das bestätigte mir vor Kurzem eine Umfrage, nach der 64 Prozent der Befragten den Lehrern bescheinigten, dass sie gute oder sogar sehr gute Arbeit leisteten.	To je nedavno potvrdila i anketa u kojoj je 64 posto ispitanika izjavilo da je radilo dobro ili čak vrlo dobro.	To je potvrdilo i nedavno istraživanje, prema kojem je 64 posto ispitanih reklo da su učitelji dobro ili čak vrlo dobro radili posao.	To mi je nedavno potvrdila i jedna anketa prema kojoj je 64 % ispitanika priznalo da nastavnici dobro ili čak vrlo dobro obavljaju svoj posao.
14	Also jubiliere und an die Brust klopfen?	Tako slaviti i kucati na grudi?	Pa navijati i kucati u prsa?	Znači, imamo se čemu veseliti?
15	Da ist aber auch die andere Seite der Medaille.	Ali postoji i druga strana medalje.	Ali postoji i druga strana medalje.	Postoji, međutim, i druga strana medalje.
16	Ich glaube, kaum ein Berufsstand wurde in den vergangenen Jahren so stark angegriffen wie der der Lehrer.	Vjerujem da je u posljednjih nekoliko godina kao učiteljica gotovo da nije bilo tako strogo napadnuto zanimanje.	Mislim da teško da postoji zanimanje koje je posljednjih godina tako ozbiljno napadnuto kao učitelji.	Mislim da nema te struke koja je proteklih godina bila tako snažno napadnuta kao nastavnička.
17	Da sind die diversen Pisa-Studien, die immer wieder belegen, dass unsere Schüler im internationalen Vergleich gerade mal nur Mittelmaß sind.	Postoje razne studije u Pisi koje dokazuju da su njemački studenti samo prosječni u međunarodnoj usporedbi.	Postoje razne studije iz Pise, koje više puta dokazuju da su naši studenti u međunarodnoj usporedbi tek osrednji.	Iskaču neke PISA-ine studije koje neprestano dokazuju da su naši učenici u međunarodnoj usporedbi tek osrednji.
18	Wer trägt dafür die Verantwortung?	Tko snosi odgovornost za to?	Tko je odgovoran za ovo?	Tko za to snosi odgovornost?
19	Letztendlich sind es immer die Lehrer, die die Sache nicht in den Griff kriegen, sich zu wenig Mühe geben oder gar für ihren Beruf nicht geeignet sind.	Uostalom, uvijek su nastavnici ti koji ne mogu kontrolirati stvari, koji ne ulažu dovoljno truda u svoj posao, ili koji nisu sposobni za svoj posao.	Na kraju, uvijek su učitelji ti koji ne mogu stvari kontrolirati, koji se ne trude dovoljno ili su čak neprikladni za svoju profesiju.	Na kraju su uvijek nastavnici ti koji navodno nemaju stvari pod kontrolom, koji se premalo trude ili čak uopće nisu sposobni za obavljanje njihova zanimanja.
20	Was können wir Lehrer tun, um die Dinge zum Besseren zu wenden?	Što mi nastavnici možemo učiniti da promijenimo stvari na bolje?	Što mi učitelji možemo učiniti da stvari promijenimo na bolje?	Što mi nastavnici možemo učiniti da okrenemo stvari nabolje?
21	Natürlich könnten wir immer wieder auf die anderen zeigen, darüber klagen, was uns angetan wird und auf unsere kaum zu bewältigenden Aufgaben verweisen – denn wir müssen nämlich oft auch	Naravno, uvijek bismo mogli ukazati na druge, žaliti se na ono što nam je učinjeno i odnositi se na teško upravljive zadatke - jer često moramo preuzeti poslove roditeljstva.	Naravno, mogli bismo uvijek iznova ukazivati na ostale, žaliti se na ono što nam se radi i pozivati se na zadatke kojima teško možemo upravljati - jer često moramo preuzeti i zadatke roditeljstva.	Naravno da bismo mogli upirati prstom u druge, žaliti se na sve ono što nam je učinjeno i ukazati na to koliko je teško savladati sve naše zadatke – jer često, naime, moramo

	Erziehungsaufgaben der Eltern übernehmen.			preuzeti i odgojnu ulogu roditelja.
22	Wir könnten uns aber auch an die eigene Nase fassen und überlegen, wie wir gegen all das angehen können.	Ali također možemo uzeti svoj nos i razmisliti o tome kako se možemo pozabaviti spomenutim Duckmäusertumom.	Ali mogli bismo pogledati i vlastiti nos i razmotriti kako se možemo suprotstaviti gore spomenutoj opreznosti.	Ali mogli bismo i kritički postaviti sami prema sebi i razmisliti o tome kako da se nosimo sa svime ovime.
23	Es ist schon komisch, dass wir Lehrer, die wir stets und ständig andere und ihre Arbeit bewerten und beurteilen, nicht in der Lage sind, souverän mit Kritik oder Urteilen über uns selber umzugehen.	Smiješno je da mi učitelji koji neprestano i stalno procjenjujemo i vrednujemo druge i njihov rad nisu u poziciji da se pouzdano pozabave kritikom ili prosudbama o nama samima.	Smiješno je da mi učitelji, koje neprestano i neprestano ocjenjujemo i njihov rad, nismo u stanju pouzdano se nositi s kritikama ili prosudbama o sebi.	Ipak je smiješno da mi nastavnici, koji neprestano (pr)ocjenjujemo druge i njihov rad, nismo u stanju suvereno se nositi s kritikom ili mišljenjima o nama samima.
24	Wenn wir Lehrer untereinander ohne Scheu und falsche Scham die Karten auf den Tisch legen, müssen wir alle zugeben, dass unser Alltag von permanenter Angst begleitet wird – der Angst, vor der Klasse zu scheitern.	Ako mi učitelji položimo jedni druge na stol bez sramežljivosti i lažne sramote, svi moramo priznati da je naš svakodnevni život popraćen stalnim strahom - strahom od neuspjeha pred razredom.	Ako mi učitelji stavljamo karte na stol jedni drugima bez sramežljivosti i lažnog srama, svi moramo priznati da je naša svakodnevica popraćena trajnim strahom - strahom od neuspjeha pred razredom.	Kada bismo si mi nastavnici bez straha i lažnog srama položili karte na stol, morali bismo priznati da nam je svakodnevica popraćena neprekidnim strahom – strahom od neuspjeha pred razredom.
25	Wir müssen zugeben, dass keine Unterrichtsstunde reine Routine ist, sondern das Gegenteil: immer wieder ein Kampf um Anerkennung, ein Ringen um Aufmerksamkeit und Disziplin.	Moramo priznati da nijedna lekcija nije čista rutina, već suprotno: uvijek borba za priznanje, borba za pažnju i disciplinu.	Moramo priznati da nijedna lekcija nije čista rutina, već upravo suprotno: uvijek iznova borba za priznanje, borba za pažnju i disciplinu.	Morali bismo priznati da nijedan nastavni sat nije čista rutina, već upravo suprotno: uvijek je on iznova borba za priznanjem, pažnjom i disciplinom.
26	Und was tun wir dagegen?	I što ćemo učiniti?	I što mi poduzimamo po tom pitanju?	A kako se borimo protiv toga?
27	Bereiten Fachlehrer gemeinsam Unterrichtseinheiten vor?	Jesu li predmetni nastavnici zajedno pripremili nastavne jedinice?	Pripremaju li učitelji predmetne nastave zajedno?	Pripremaju li predmetni nastavnici zajednički lekcije?
28	Probieren wir alternative Methoden aus?	Pokušavamo li alternativne metode?	Pokušavamo li alternativne metode?	Iskušavamo li alternativne metode?
29	Entwickeln wir gemeinsam Kommunikationsformen, die uns näher an die Wünsche und Vorstellungen unserer Schüler bringen?	Da li zajedno razvijamo oblike komunikacije koji nas približavaju željama i idejama naših učenika?	Razvijamo li zajednički oblike komunikacije koji nas približavaju željama i idejama naših učenika?	Razvijamo li zajednički oblike komunikacije koji bi nas približili željama i predodžbama naših učenika?
30	Leider nicht.	Nažalost ne.	Nažalost ne.	Nažalost, ne.

31	In der Mehrzahl der Fälle bleibt jeder Kollege der traditionelle Einzelkämpfer und das, was sich hinter der Klassentür abspielt, bleibt geheim.	U većini slučajeva, svaki kolega ostaje tradicionalni usamljeni borac, a ono što se događa iza učionice ostaje tajna.	U većini slučajeva svaki kolega ostaje tradicionalni usamljeni ratnik, a ono što se događa iza vrata razreda ostaje tajna.	U većini slučajeva svaki kolega bori se sam za sebe, a ono što se odvija iza zatvorenih vrata učionice ostaje tajna.
32	Es gibt inzwischen Schulen, an denen eine kontinuierliche Lehrerbewertung durch die Schüler eingeführt wurde.	Sada postoje škole u kojima su studenti provodili kontinuirano ocjenjivanje nastavnika.	Sada postoje škole u kojima je uvedeno kontinuirano ocjenjivanje nastavnika od strane učenika.	U međuvremenu postoje škole u kojima je uvedeno kontinuirano vrednovanje nastavnika koje provode učenici.
33	Damit sie funktioniert, muss dreierlei passieren: Erstens müssen die bewertenden Schüler anonym bleiben.	Postoje tri stvari koje ga čine uspješnim: Prvo, učenici koji ocjenjuju moraju ostati anonimni.	Da bi to uspjelo, moraju se dogoditi tri stvari: Prvo, studenti koji ocjenjuju moraju ostati anonimni.	Da bi takvo vrednovanje funkcioniralo, moraju se dogoditi tri stvari: Učenici koji vrednuju nastavnike prvenstveno moraju ostati anonimni.
34	Die Schüler müssen zweitens gelernt haben, wie man Kritik übt.	Drugo, učenici su morali naučiti kritizirati.	Drugo, studenti su sigurno naučili kako kritizirati.	Učenici zatim moraju naučiti kako izreći kritiku.
35	Und drittens erfährt allein die betroffene Lehrkraft die Ergebnisse, mit denen sie dann umgehen kann, wie sie will.	I treće, samo učitelj o kojem se radi uči rezultate, koje oni mogu podnijeti kako žele.	I treće, samo dotični učitelj doživljava rezultate s kojima se onda može nositi kako želi.	I, treće, samo dotični nastavnik smije saznati rezultate vrednovanja te s njima potom smije raditi što želi.
36	Bisher ist erst ein Viertel der bewerteten Lehrerinnen und Lehrer bereit, mit ihrer jeweiligen Klasse über die geäußerte Kritik zu sprechen.	Do sada je samo četvrtina ispitanih nastavnika spremna razgovarati sa svojim razredom o kritikama.	Do sada je samo četvrtina ocijenjenih učitelja spremna razgovarati sa svojim razredom o izrečenim kritikama.	Dosad je tek četvrtina vrednovanih nastavnika i nastavnica spremna sa svojim razredom porazgovarati o izrečenoj kritici.
37	Manchen Schülern ist das zu wenig und sie fragen nach dem Sinn der ganzen Frageaktion.	Ovo nije dovoljno za neke učenike i oni pitaju za značenje cijele reakcije pitanja.	Nekim studentima to nije dovoljno i pitaju o značenju cijelog postupka ispitivanja.	Neki učenici smatraju da to nije dovoljno i pitaju koji je uopće smisao čitavog upitnika.
38	Ihnen möchte ich antworten: Habt Geduld, immerhin ist ein Anfang gemacht.	Želim odgovoriti: Imajte strpljenja, nakon svega, napravljen je početak.	Želio bih vam odgovoriti: budite strpljivi, barem je početak započet.	Njima bih htjela odgovoriti: Strpite se, barem se nešto pokrenulo.

Prilog B

Anketni obrazac:

Ispunjavanjem ovog anketnog upitnika sudjelujete u istraživanju vrednovanja razumljivosti strojnih prijevoda. Istraživanje je anonimno; vaši su podaci u potpunosti zaštićeni i rabić će se isključivo u znanstvene svrhe. Želite li odustati od ispunjavanja upitnika, svi će se vaši odgovori i podaci automatski obrisati. Hvala na izdvojenom vremenu!

Prvi zadatak (1/2) (prva skupina ispitanika)

Ispod ovog okvira nalazi se trideset i osam strojno prevedenih rečenica, izvučenih iz jedne nastavničke kolumne i nasumično poredanih. Pročitajte rečenice i ocijenite koliko su vam razumljive na ljestvici od 1 do 5, prema sljedećim vrijednostima:

- 1 rečenica je nerazumljiva
- 2 rečenica je slabo razumljiva
- 3 rečenica je prilično razumljiva
- 4 rečenica je dobro razumljiva
- 5 rečenica je potpuno razumljiva

U ovom zadatku nema točnih i netočnih odgovora jer se radi isključivo o vašoj osobnoj procjeni prijevoda.

Postoje razne studije iz Pise, koje više puta dokazuju da su naši studenti u međunarodnoj usporedbi tek osrednji. *

1 2 3 4 5

Ali postoji i druga strana medalje. *

1 2 3 4 5

Ako mi učitelji stavljamo karte na stol jedni drugima bez sramežljivosti i lažnog srama, svi moramo priznati da je naša svakodnevnica popraćena trajnim strahom - strahom od neuspjeha pred razredom. *

1 2 3 4 5

Pokušavamo li alternativne metode? *	1	2	3	4	5
Rad s djecom i mladima, prenošenje znanja i vrijednosti - nisam mogao smisliti bolji posao. *	1	2	3	4	5
Ova prva rečenica nekako nije samo kratka i jezgrovita informacija o mom poslu, već priznanje. *	1	2	3	4	5
Moramo priznati da nijedna lekcija nije čista rutina, već upravo suprotno: uvijek iznova borba za priznanje, borba za pažnju i disciplinu. *	1	2	3	4	5
Nažalost ne. *	1	2	3	4	5
Razvijamo li zajednički oblike komunikacije koji nas približavaju željama i idejama naših učenika? *	1	2	3	4	5
Sada postoje škole u kojima je uvedeno kontinuirano ocjenjivanje nastavnika od strane učenika. *	1	2	3	4	5
Zašto? *	1	2	3	4	5
I što mi poduzimamo po tom pitanju? *	1	2	3	4	5
Da bi to uspjelo, moraju se dogoditi tri stvari: Prvo, studenti koji ocjenjuju moraju ostati anonimni. *	1	2	3	4	5
Nekim studentima to nije dovoljno i pitaju o značenju cijelog postupka ispitivanja. *	1	2	3	4	5
Drugo, studenti su sigurno naučili kako kritizirati. *	1	2	3	4	5
Tko je odgovoran za ovo? *	1	2	3	4	5

U većini slučajeva svaki kolega ostaje tradicionalni usamljeni ratnik, a ono što se događa iza vrata razreda ostaje tajna. *	1	2	3	4	5
Rečenica "Ja sam učitelj" uvijek mi se čini kao priznanje, kao priznanje nečemu negativnom, otprilike slično priznanju poroku poput pušenja ili još gore. *	1	2	3	4	5
Ne zbog statusa državnog službenika, koji me ne izlaže neizvjesnostima na tržištu rada. *	1	2	3	4	5
Na kraju, uvijek su učitelji ti koji ne mogu stvari kontrolirati, koji se ne trude dovoljno ili su čak neprikladni za svoju profesiju. *	1	2	3	4	5
Želio bih vam odgovoriti: budite strpljivi, barem je početak započet. *	1	2	3	4	5
Naravno, mogli bismo uvijek iznova ukazivati na ostale, žaliti se na ono što nam se radi i pozivati se na zadatke kojima teško možemo upravljati - jer često moramo preuzeti i zadatke roditeljstva. *	1	2	3	4	5
Smiješno je da mi učitelji, koje neprestano i neprestano ocjenjujemo i ocjenjujemo druge i njihov rad, nismo u stanju pouzdano se nositi s kritikama ili prosudbama o sebi. *	1	2	3	4	5
Ali mogli bismo pogledati i vlastiti nos i razmotriti kako se svemu tome možemo suprotstaviti. *	1	2	3	4	5
Do sada je samo četvrtina ocijenjenih učitelja spremna razgovarati sa svojim razredom o izrečenim kritikama. *	1	2	3	4	5

To je potvrdilo i nedavno istraživanje, prema kojem je 64 posto ispitanih reklo da su učitelji dobro ili čak vrlo dobro radili posao. *	1	2	3	4	5
Pa navijati i kucati u prsa? *	1	2	3	4	5
Ugodno je, priznat ću, ali nije presudno. *	1	2	3	4	5
Mislim da teško da postoji zanimanje koje je posljednjih godina tako ozbiljno napadnuto kao učitelji. *	1	2	3	4	5
„Nemojte se samo žaliti!“ *	1	2	3	4	5
I treće, samo dotični učitelj doživljava rezultate s kojima se onda može nositi kako želi. *	1	2	3	4	5
I svo slobodno vrijeme - to je ionako samo iluzija. *	1	2	3	4	5
Volim svoj posao. *	1	2	3	4	5
Što mi učitelji možemo učiniti da stvari promijenimo na bolje? *	1	2	3	4	5
Ja sam učiteljica i predajem njemački jezik i povijest. *	1	2	3	4	5
Pripremaju li učitelji predmetne nastave zajedno? *	1	2	3	4	5
Ne, kad sam studirao "za učitelja", bio sam pun ideala. *	1	2	3	4	5
Zapravo, ne bih se trebao skrivati sa svojim poslom, mogao bih biti ponosan na njega. *	1	2	3	4	5

Prvi zadatak (1/2) (druga skupina ispitanika)

Ispod ovog okvira nalazi se jedna strojno prevedena nastavnička kolumna. Pročitajte je i ocijenite koliko vam je tekst razumljiv na ljestvici od 1 do 5. U ovom zadatku nema točnog i netočnog odgovora jer se radi isključivo o vašoj osobnoj procjeni prijevoda.

„Nemojte se samo žaliti!”

Ja sam učiteljica i predajem njemački jezik i povijest.

Ova prva rečenica nekako nije samo kratka i jezgrovita informacija o mom poslu, već priznanje. Rečenica "Ja sam učitelj" uvijek mi se čini kao priznanje, kao priznanje nečemu negativnom, otprilike slično priznanju poroku poput pušenja ili još gore.

Zašto? Volim svoj posao. Ne zbog statusa državnog službenika, koji me ne izlaže neizvjesnostima na tržištu rada. Ugodno je, priznat ću, ali nije presudno. I svo slobodno vrijeme - to je ionako samo iluzija. Ne, kad sam studirao "za učitelja", bio sam pun ideala. Rad s djecom i mladima, prenošenje znanja i vrijednosti - nisam mogao smisliti bolji posao.

Zapravo, ne bih se trebao skrivati sa svojim poslom, mogao bih biti ponosan na njega. To je potvrdilo i nedavno istraživanje, prema kojem je 64 posto ispitanih reklo da su učitelji dobro ili čak vrlo dobro radili posao. Pa navijati i kucati u prsa?

Ali postoji i druga strana medalje. Mislim da teško da postoji zanimanje koje je posljednjih godina tako ozbiljno napadnuto kao učitelji. Postoje razne studije iz Pise, koje više puta dokazuju da su naši studenti u međunarodnoj usporedbi tek osrednji. Tko je odgovoran za ovo? Na kraju, uvijek su učitelji ti koji ne mogu stvari kontrolirati, koji se ne trude dovoljno ili su čak neprikladni za svoju profesiju.

Što mi učitelji možemo učiniti da stvari promijenimo na bolje? Naravno, mogli bismo uvijek iznova ukazivati na ostale, žaliti se na ono što nam se radi i pozivati se na zadatke kojima teško možemo upravljati - jer često moramo preuzeti i zadatke roditeljstva.

Ali mogli bismo pogledati i vlastiti nos i razmotriti kako se svemu tome možemo suprotstaviti. Smiješno je da mi učitelji, koje neprestano i neprestano ocjenjujemo i ocjenjujemo druge i

njihov rad, nismo u stanju pouzdano se nositi s kritikama ili prosudbama o sebi.

Ako mi učitelji stavljamo karte na stol jedni drugima bez sramežljivosti i lažnog srama, svi moramo priznati da je naša svakodnevica popraćena trajnim strahom - strahom od neuspjeha pred razredom. Moramo priznati da nijedna lekcija nije čista rutina, već upravo suprotno: uvijek iznova borba za priznanje, borba za pažnju i disciplinu.

I što mi poduzimamo po tom pitanju? Pripremaju li učitelji predmetne nastave zajedno? Pokušavamo li alternativne metode? Razvijamo li zajednički oblike komunikacije koji nas približavaju željama i idejama naših učenika? Nažalost ne. U većini slučajeva svaki kolega ostaje tradicionalni usamljeni ratnik, a ono što se događa iza vrata razreda ostaje tajna.

Sada postoje škole u kojima je uvedeno kontinuirano ocjenjivanje nastavnika od strane učenika. Da bi to uspjelo, moraju se dogoditi tri stvari: Prvo, studenti koji ocjenjuju moraju ostati anonimni. Drugo, studenti su sigurno naučili kako kritizirati. I treće, samo dotični učitelj doživljava rezultate s kojima se onda može nositi kako želi.

Do sada je samo četvrtina ocijenjenih učitelja spremna razgovarati sa svojim razredom o izrečenim kritikama. Nekim studentima to nije dovoljno i pitaju o značenju cijelog postupka ispitivanja. Želio bih vam odgovoriti: budite strpljivi, barem je početak započet.

Ovaj prijevod ocjenjujem ocjenom: *

1 – nerazumljiv je

2 – slabo je razumljiv

3 – prilično je razumljiv

4 – dobro je razumljiv

5 – potpuno je razumljiv

Druzi zadatak (2/2) (obje skupine ispitanika)

Pred vama je ponovno ista strojno prevedena kolumna. Ovisno o verziji istraživanja koju ste dobili možda prvi put vidite kolumnu kao cjelovit tekst. Pročitajte je, a potom odgovorite na sedam pitanja o tekstu. Samo je jedan odgovor na svako pitanje točan. Pri odgovaranju na pitanja slobodno se po potrebi vratite na tekst.

VAŽNO! Rješavanje ovog zadatka nije vremenski ograničeno. Kako biste što točnije odgovorili na pitanja, uzmite koliko god vremena želite. Na kraju zadatka unesite potrebno vrijeme rješavanja izraženo u minutama. Molimo da počnete računati vrijeme od sad.

„Nemojte se samo žaliti!”

Ja sam učiteljica i predajem njemački jezik i povijest.

Ova prva rečenica nekako nije samo kratka i jezgrovita informacija o mom poslu, već priznanje. Rečenica "Ja sam učitelj" uvijek mi se čini kao priznanje, kao priznanje nečemu negativnom, otprilike slično priznanju poroku poput pušenja ili još gore.

Zašto? Volim svoj posao. Ne zbog statusa državnog službenika, koji me ne izlaže neizvjesnostima na tržištu rada. Ugodno je, priznat ću, ali nije presudno. I svo slobodno vrijeme - to je ionako samo iluzija. Ne, kad sam studirao "za učitelja", bio sam pun ideala. Rad s djecom i mladima, prenošenje znanja i vrijednosti - nisam mogao smisliti bolji posao.

Zapravo, ne bih se trebao skrivati sa svojim poslom, mogao bih biti ponosan na njega. To je potvrdilo i nedavno istraživanje, prema kojem je 64 posto ispitanih reklo da su učitelji dobro ili čak vrlo dobro radili posao. Pa navijati i kucati u prsa?

Ali postoji i druga strana medalje. Mislim da teško da postoji zanimanje koje je posljednjih godina tako ozbiljno napadnuto kao učitelji. Postoje razne studije iz Pise, koje više puta dokazuju da su naši studenti u međunarodnoj usporedbi tek osrednji. Tko je odgovoran za ovo? Na kraju, uvijek su učitelji ti koji ne mogu stvari kontrolirati, koji se ne trude dovoljno ili su čak neprikladni za svoju profesiju.

Što mi učitelji možemo učiniti da stvari promijenimo na bolje? Naravno, mogli bismo uvijek

iznova ukazivati na ostale, žaliti se na ono što nam se radi i pozivati se na zadatke kojima teško možemo upravljati - jer često moramo preuzeti i zadatke roditeljstva.

Ali mogli bismo pogledati i vlastiti nos i razmotriti kako se svemu tome možemo suprotstaviti. Smiješno je da mi učitelji, koje neprestano i neprestano ocjenjujemo i ocjenjujemo druge i njihov rad, nismo u stanju pouzdano se nositi s kritikama ili prosudbama o sebi.

Ako mi učitelji stavljamo karte na stol jedni drugima bez sramežljivosti i lažnog srama, svi moramo priznati da je naša svakodnevica popraćena trajnim strahom - strahom od neuspjeha pred razredom. Moramo priznati da nijedna lekcija nije čista rutina, već upravo suprotno: uvijek iznova borba za priznanje, borba za pažnju i disciplinu.

I što mi poduzimamo po tom pitanju? Pripremaju li učitelji predmetne nastave zajedno? Pokušavamo li alternativne metode? Razvijamo li zajednički oblike komunikacije koji nas približavaju željama i idejama naših učenika? Nažalost ne. U većini slučajeva svaki kolega ostaje tradicionalni usamljeni ratnik, a ono što se događa iza vrata razreda ostaje tajna.

Sada postoje škole u kojima je uvedeno kontinuirano ocjenjivanje nastavnika od strane učenika. Da bi to uspjelo, moraju se dogoditi tri stvari: Prvo, studenti koji ocjenjuju moraju ostati anonimni. Drugo, studenti su sigurno naučili kako kritizirati. I treće, samo dotični učitelj doživljava rezultate s kojima se onda može nositi kako želi.

Do sada je samo četvrtina ocijenjenih učitelja spremna razgovarati sa svojim razredom o izrečenim kritikama. Nekim studentima to nije dovoljno i pitaju o značenju cijelog postupka ispitivanja. Želio bih vam odgovoriti: budite strpljivi, barem je početak započet.

1) Kako se osoba osjeća kada govori o svom zanimanju? *

Osjeća se vrlo važno.

Ima pomiješane osjećaje.

Ljuti se na one koji je to pitaju.

Ne osjeća ništa.

2) Zašto se osoba odlučila za nastavničko zanimanje? *

- Htjela je mlade podučavati stvarima važnima za život.
- Tražila je posao koji joj pruža mnogo slobode.
- Sviđala joj se sigurnost tog radnog mjesta.
- Tražila je odgovoran posao.

3) Kakav je, prema mišljenju ove osobe, stav javnosti prema nastavničkom zanimanju? *

- Nastavnike se krivi za sve što je u društvu loše.
- Negativan stav prema nastavnicima više nije toliko izražen kao prije.
- Nastavnici su istovremeno i vrlo priznati i snažno ih se kritizira.
- Nastavnike se uglavnom smatra nesposobnima.

4) S obzirom na opisanu situaciju, ova se osoba zalaže za to da... *

- ...roditelji preuzmu veću odgovornost za odgoj svoje djece.
- ...društvo pruži bolju podršku nastavnicima.
- ...se učenicima smanji pritisak i opterećenost uspjesima.
- ...nastavnici razmisle o tome kako sami mogu promijeniti neke stvari nabolje.

5) Osoba smatra da se nastavnici boje... *

- ...da nisu u tijeku s najnovijim znanstvenim i stručnim spoznajama.
- ...da ih učenici neće dobro prihvatiti.
- ...razgovarati s kolegama o svojim problemima.
- ...u nastavi iskušati nešto novo.

6) Osoba smatra da bi bilo dobro kada bi... *

- ...nastavnici zajedničkim snagama osvježavali didaktičke pristupe u nastavi.
- ...nastavnici snažno usmjerili pažnju javnosti na svoje probleme.
- ...kolege vrednovali rad svakog nastavnika.
- ...se u suradnji s učenicima razvijali novi modeli podučavanja i učenja.

7) Prva iskustva s vrednovanjem nastavnika pokazuju da... *

- ...su tek rijetki učenici spremni pružiti konstruktivnu kritiku.
- ...učenici brzo izgube volju za sudjelovanjem.
- ...se učenici boje da će se ugroziti njihovo pravo na anonimnost.
- ...tek rijetki nastavnici razgovaraju o kritici koju su dobili od učenika.

Procijenite koliko ste sigurni u svoje odgovore na pitanja. *

- Nimalo nisam siguran/na u odabir odgovora.
- Pretpostavljam da sam odabrao/la većinu netočnih odgovora.
- Niti sam siguran/na niti nisam.
- Pretpostavljam da sam odabrao/la većinu točnih odgovora.
- Sasvim sam siguran/na u odabir odgovora.

Vrijeme potrebno za ispunjavanje obrasca (u minutama): *

Hvala na rješavanju zadataka!

Prije podnošenja odgovora molimo da na ljestvici od 1 do 5 ocijenite koliko se slažete sa sljedećih šest izjava.

U profesionalnom i/ili privatnom životu aktivno se koristim alatima za strojno prevođenje (npr. Google Prevoditelj, Bing Translate i sl.).*

uopće se ne slažem 1 2 3 4 5 potpuno se slažem

Automatski prijevodi znaju mi poslužiti za razumijevanje, primjerice, statusa na stranom jeziku na društvenim mrežama ili neke mrežne stranice. *

uopće se ne slažem 1 2 3 4 5 potpuno se slažem

Smatram da su alati za strojno prevođenje ljudima korisni u svakodnevnim situacijama. *

uopće se ne slažem 1 2 3 4 5 potpuno se slažem

Ne sviđa mi se ideja razvitka strojnog prevođenja i otimanje posla stručnim ljudima. *

uopće se ne slažem 1 2 3 4 5 potpuno se slažem

Strojni prijevodi i sada već postižu mnogo i veselim se pozitivnom napretku u budućnosti. *

uopće se ne slažem 1 2 3 4 5 potpuno se slažem

Ako želite, možete i vlastitim riječima podijeliti s nama vaše mišljenje o strojnom prevođenju, korisnosti alata za strojno prevođenje, razumljivosti strojnih prijevoda i sl.

Demografski podaci:

Spol M
 Ž

Dob * 18-25
 26-35
 36-45
 46-55
 56-65
 više od 65

Zadnji završeni stupanj obrazovanja * osnovna škola
 trogodišnja srednja škola
 četverogodišnja srednja škola
 stručni ili sveučilišni preddiplomski studij
 stručni ili sveučilišni diplomski studij / integrirani studij
 doktorski studij

Zanimanje *

*Ako ste po zanimanju prevoditelj, molimo vas da navedete vašu glavnu užu specijalizaciju (primjerice: tehnički prevoditelj, stručni prevoditelj, usmeni prevoditelj, sudski tumač...)

Vaši su odgovori zabilježeni! Hvala što ste sudjelovali u istraživanju!

Prilog C

Razina slaganja ispitanika s navedenim izjavama (od 1 – uopće se ne slažem do 5 – u potpunosti se slažem) te kodovi stavova prema stupnju slaganja (pozitivan, neutralan ili negativan stav) s ukupnim određenjem stava za ispitanika (+ pozitivan, / neutralan, - negativan):

	U profesionalnom i/ili privatnom životu aktivno se koristim alatima za strojno prevođenje (npr. Google Prevoditelj, Bing Translate i sl.).	Automatski prijevodi znaju mi poslužiti za razumijevanje, primjerice, statusa na stranom jeziku na društvenim mrežama ili neke mrežne stranice.	Smatram da su alati za strojno prevođenje ljudima korisni u svakodnevnim situacijama.	Ne sviđa mi se ideja razvitka strojnog prevođenja i otimanje posla stručnim ljudima.	Strojni prijevodi i sada već postižu mnogo i veselim se pozitivnom napretku u budućnosti.	
I1	3, neutralan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	/
I2	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I3	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I4	2, negativan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I5	4, pozitivan stav.	2, negativan stav.	4, pozitivan stav.	5, negativan stav.	2, negativan stav.	-
I6	2, negativan stav.	4, pozitivan stav.	4, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I7	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I8	3, neutralan stav.	2, negativan stav.	4, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I9	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I10	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	3, neutralan stav.	2, negativan stav.	/
I11	3, neutralan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I12	4, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	4, negativan stav.	4, pozitivan stav.	+
I13	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I14	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	+
I15	2, negativan stav.	5, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	4, pozitivan stav.	+
I16	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I17	2, negativan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+

I18	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I19	2, negativan stav.	1, negativan stav.	3, neutralan stav.	1, pozitivan stav.	2, negativan stav.	-
I20	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	4, pozitivan stav.	+
I21	4, pozitivan stav.	2, negativan stav.	5, pozitivan stav.	5, negativan stav.	3, neutralan stav.	/
I22	2, negativan stav.	2, negativan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	/
I23	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	5, negativan stav.	3, neutralan stav.	+
I24	2, negativan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I25	3, neutralan stav.	1, negativan stav.	3, neutralan stav.	5, negativan stav.	3, neutralan stav.	/
I26	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	3, neutralan stav.	+
I27	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I28	1, negativan stav.	3, neutralan stav.	3, neutralan stav.	2, pozitivan stav.	4, pozitivan stav.	/
I29	4, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	+
I30	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I31	4, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I32	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I33	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	4, pozitivan stav.	+
I34	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I35	3, neutralan stav.	4, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I36	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	5, negativan stav.	2, negativan stav.	/
I37	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I38	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I39	4, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	5, negativan stav.	5, pozitivan stav.	+
I40	3, neutralan stav.	3, neutralan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	/
I41	3, neutralan stav.	2, negativan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I42	3, neutralan stav.	2, negativan stav.	3, neutralan stav.	4, negativan stav.	3, neutralan stav.	/
I43	2, negativan stav.	4, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I44	5, pozitivan stav.	5, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	+
I45	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+

I46	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I47	2, negativan stav.	2, negativan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I48	1, negativan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	4, pozitivan stav.	/
I49	3, neutralan stav.	3, neutralan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I50	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	3, neutralan stav.	+
I51	2, negativan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	/
I52	3, neutralan stav.	3, neutralan stav.	4, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	/
I53	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I54	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I55	4, pozitivan stav.	5, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I56	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I57	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I58	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I59	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I60	4, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I61	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I62	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	5, pozitivan stav.	+
I63	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I64	4, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	5, negativan stav.	4, pozitivan stav.	+
I65	3, neutralan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I66	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I67	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	5, pozitivan stav.	+
I68	2, negativan stav.	1, negativan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I69	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I70	2, negativan stav.	3, neutralan stav.	3, neutralan stav.	5, negativan stav.	3, neutralan stav.	/
I71	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	4, pozitivan stav.	+
I72	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I73	3, neutralan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+

I74	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	5, pozitivan stav.	+
I75	4, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I76	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I77	5, pozitivan stav.	5, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I78	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I79	5, pozitivan stav.	5, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I80	2, negativan stav.	3, neutralan stav.	3, neutralan stav.	2, pozitivan stav.	3, neutralan stav.	/
I81	3, neutralan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I82	1, negativan stav.	1, negativan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	/
I83	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I84	4, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	5, negativan stav.	3, neutralan stav.	+
I85	3, neutralan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I86	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I87	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I88	3, neutralan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I89	2, negativan stav.	2, negativan stav.	3, neutralan stav.	1, pozitivan stav.	4, pozitivan stav.	/
I90	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	5, negativan stav.	3, neutralan stav.	/
I91	1, negativan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I92	4, pozitivan stav.	2, negativan stav.	5, pozitivan stav.	5, negativan stav.	2, negativan stav.	-
I93	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I94	3, neutralan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I95	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I96	3, neutralan stav.	2, negativan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I97	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I98	3, neutralan stav.	3, neutralan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	/
I99	5, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	1, pozitivan stav.	3, neutralan stav.	+
I100	2, negativan stav.	4, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	/
I101	4, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	/

I102	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	3, neutralan stav.	+
I103	2, negativan stav.	2, negativan stav.	3, neutralan stav.	4, negativan stav.	1, negativan stav.	-
I104	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I105	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I106	3, neutralan stav.	2, negativan stav.	3, neutralan stav.	4, negativan stav.	3, neutralan stav.	/
I107	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	4, negativan stav.	4, pozitivan stav.	+
I108	3, neutralan stav.	3, neutralan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	/
I109	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	2, negativan stav.	/
I110	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	4, negativan stav.	3, neutralan stav.	/
I111	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I112	3, neutralan stav.	2, negativan stav.	5, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I113	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I114	4, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I115	4, pozitivan stav.	2, negativan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I116	5, pozitivan stav.	2, negativan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I117	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I118	4, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I119	3, neutralan stav.	4, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	2, negativan stav.	+
I120	3, neutralan stav.	1, negativan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I121	2, negativan stav.	2, negativan stav.	4, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	/
I122	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I123	4, pozitivan stav.	1, negativan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I124	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	1, pozitivan stav.	3, neutralan stav.	+
I125	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	4, negativan stav.	3, neutralan stav.	/
I126	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, negativan stav.	3, neutralan stav.	/
I127	4, pozitivan stav.	5, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I128	3, neutralan stav.	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I129	5, pozitivan stav.	2, negativan stav.	3, neutralan stav.	1, pozitivan stav.	5, pozitivan stav.	+

I130	4, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I131	5, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	5, negativan stav.	4, pozitivan stav.	+
I132	2, negativan stav.	2, negativan stav.	4, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I133	4, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I134	2, negativan stav.	1, negativan stav.	4, pozitivan stav.	4, negativan stav.	5, pozitivan stav.	-
I135	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I136	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
I137	4, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I138	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I139	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I140	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
I141	4, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I142	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I143	3, neutralan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I144	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	4, negativan stav.	3, neutralan stav.	/
I145	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I146	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	+
I147	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
I148	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I149	4, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	+
I150	5, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I151	3, neutralan stav.	5, pozitivan stav.	4, pozitivan stav.	1, pozitivan stav.	3, neutralan stav.	+
I152	4, pozitivan stav.	5, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
I153	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	/
I154	5, pozitivan stav.	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
I155	4, pozitivan stav.	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	+
I156	2, negativan stav.	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
I157	3, neutralan stav.	1, negativan stav.	3, neutralan stav.	3, neutralan stav.	5, pozitivan stav.	/

I158	4, pozitivan stav.	4, pozitivan stav.	4, pozitivan stav.	4, negativan stav.	3, neutralan stav.	+
I159	3, neutralan stav.	5, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	/
I160	4, pozitivan stav.	5, pozitivan stav.	4, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+

Razina slaganja ispitanika predistraživanja (2019.) s navedenim izjavama (od 1 – uopće se ne slažem do 5 – u potpunosti se slažem) te kodovi stavova prema stupnju slaganja (pozitivan, neutralan ili negativan stav) s ukupnim određenjem stava za ispitanika (+ pozitivan, / neutralan, - negativan):

	U profesionalnom i/ili privatnom životu aktivno se koristim alatima za strojno prevođenje (npr. Google Prevoditelj, Bing Translate i sl.).	Alate za strojno prevođenje smatram korisnima.	Strojevi i računalni programi prevoditeljima oduzimaju posao.	Strojevi i računalni programi prevoditeljima olakšavaju prevođenje.	
i1	5, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	5, pozitivan stav.	+
i2	4, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
i3	4, pozitivan stav.	2, negativan stav.	1, pozitivan stav.	4, pozitivan stav.	+
i4	4, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	5, pozitivan stav.	+
i5	4, pozitivan stav.	3, neutralan stav.	5, negativan stav.	3, neutralan stav.	/
i6	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	1, negativan stav.	+
i7	5, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
i8	1, negativan stav.	2, negativan stav.	5, negativan stav.	3, neutralan stav.	-
i9	4, pozitivan stav.	4, pozitivan stav.	3, neutralan stav.	4, pozitivan stav.	+
i10	3, neutralan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
i11	2, negativan stav.	4, pozitivan stav.	4, negativan stav.	3, neutralan stav.	-
i12	3, neutralan stav.	5, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
i13	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
i14	3, neutralan stav.	2, negativan stav.	2, pozitivan stav.	2, negativan stav.	-
i15	3, neutralan stav.	4, pozitivan stav.	3, neutralan stav.	2, negativan stav.	/
i16	5, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
i17	4, pozitivan stav.	4, pozitivan stav.	4, negativan stav.	4, pozitivan stav.	+
i18	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	3, neutralan stav.	+
i19	3, neutralan stav.	2, negativan stav.	3, neutralan stav.	4, pozitivan stav.	/

i20	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
i21	3, neutralan stav.	4, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
i22	2, negativan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	/
i23	4, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
i24	4, pozitivan stav.	3, neutralan stav.	4, negativan stav.	2, negativan stav.	-
i25	2, negativan stav.	2, negativan stav.	1, pozitivan stav.	4, pozitivan stav.	/
i26	3, neutralan stav.	4, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
i27	3, neutralan stav.	4, pozitivan stav.	2, pozitivan stav.	3, neutralan stav.	+
i28	4, pozitivan stav.	4, pozitivan stav.	2, pozitivan stav.	5, pozitivan stav.	+
i29	2, negativan stav.	4, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+
i30	2, negativan stav.	3, neutralan stav.	2, pozitivan stav.	4, pozitivan stav.	+
i31	3, neutralan stav.	4, pozitivan stav.	2, pozitivan stav.	3, neutralan stav.	+
i32	2, negativan stav.	3, neutralan stav.	2, pozitivan stav.	4, pozitivan stav.	+
i33	5, pozitivan stav.	3, neutralan stav.	4, negativan stav.	3, neutralan stav.	/
i34	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	4, pozitivan stav.	+
i35	4, pozitivan stav.	3, neutralan stav.	2, pozitivan stav.	4, pozitivan stav.	+
i36	5, pozitivan stav.	5, pozitivan stav.	4, negativan stav.	5, pozitivan stav.	+
i37	4, pozitivan stav.	4, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
i38	5, pozitivan stav.	5, pozitivan stav.	1, pozitivan stav.	4, pozitivan stav.	+
i39	5, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	4, pozitivan stav.	+
i40	4, pozitivan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	/
i41	2, negativan stav.	2, negativan stav.	1, pozitivan stav.	5, pozitivan stav.	+
i42	2, negativan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	/
i43	2, negativan stav.	3, neutralan stav.	3, neutralan stav.	3, neutralan stav.	/
i44	5, pozitivan stav.	5, pozitivan stav.	3, neutralan stav.	5, pozitivan stav.	+
i45	5, pozitivan stav.	5, pozitivan stav.	2, pozitivan stav.	4, pozitivan stav.	+

Prilog D

Rezultati testa razumijevanja (2021.):

	Kako se osoba osjeća kada govori o svom zanimanju?	Zašto se osoba odlučila za nastavničko zanimanje?	Kakav je, prema mišljenju ove osobe, stav javnosti prema nastavničkom zanimanju?	S obzirom na opisanu situaciju, ova se osoba zalaže za to da...	Osoba smatra da se nastavnici boje...	Osoba smatra da bi bilo dobro kada bi...	Prva iskustva s vrednovanjem nastavnika pokazuju da...
I1	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I2	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Točno: A	Točno: D
I3	Točno: B	Netočno: B	Netočno: A	Netočno: B	Netočno: D	Netočno: B	Netočno: A
I4	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I5	Netočno: A	Točno: A	Netočno: D	Točno: D	Točno: B	Netočno: D	Točno: D
I6	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Točno: A	Točno: D
I7	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Točno: A	Točno: D
I8	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I9	Točno: B	Točno: A	Netočno: A	Netočno: B	Netočno: C	Netočno: D	Točno: D
I10	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Netočno: D	Točno: D
I11	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Netočno: D	Točno: D
I12	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I13	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Netočno: D	Točno: D
I14	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I15	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Točno: A	Točno: D
I16	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Netočno: C	Točno: D
I17	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Točno: A	Točno: D
I18	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Netočno: D	Točno: D
I19	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
I20	Netočno: C	Točno: A	Netočno: B	Točno: D	Netočno: C	Točno: A	Točno: D
I21	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: B	Točno: D

I22	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Netočno: D	Točno: D
I23	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Točno: A	Točno: D
I24	Točno: B	Točno: A	Točno: C	Netočno: B	Netočno: D	Točno: A	Točno: D
I25	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Točno: A	Točno: D
I26	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: C	Točno: D
I27	Točno: B	Točno: A	Netočno: A	Netočno: A	Točno: B	Netočno: B	Netočno: B
I28	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Netočno: D	Točno: D
I29	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
I30	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Netočno: D	Točno: D
I31	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I32	Netočno: A	Točno: A	Netočno: A	Točno: D	Netočno: C	Netočno: D	Točno: D
I33	Točno: B	Točno: A	Točno: C	Netočno: B	Netočno: C	Točno: A	Točno: D
I34	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Točno: A	Točno: D
I35	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
I36	Točno: B	Točno: A	Netočno: A	Netočno: B	Netočno: C	Netočno: C	Točno: D
I37	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I38	Netočno: A	Netočno: B	Točno: C	Točno: D	Točno: B	Točno: A	Netočno: A
I39	Netočno: A	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: B	Netočno: C
I40	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Netočno: D	Točno: D
I41	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Točno: A	Točno: D
I42	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Netočno: D	Točno: D
I43	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Točno: A	Točno: D
I44	Točno: B	Točno: A	Netočno: B	Netočno: A	Točno: B	Netočno: D	Netočno: C
I45	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
I46	Točno: B	Točno: A	Netočno: A	Netočno: C	Netočno: C	Netočno: D	Točno: D
I47	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Točno: D
I48	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I49	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Netočno: D	Točno: D
I50	Točno: B	Točno: A	Netočno: D	Netočno: B	Netočno: D	Netočno: D	Točno: D
I51	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I52	Točno: B	Točno: A	Točno: C	Netočno: B	Netočno: C	Netočno: D	Točno: D
I53	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Netočno: B	Netočno: A

I54	Netočno: A	Točno: A	Netočno: D	Točno: D	Netočno: D	Netočno: D	Točno: D
I55	Točno: B	Netočno: D	Točno: C	Netočno: B	Točno: B	Netočno: B	Netočno: B
I56	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Točno: D
I57	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Netočno: D	Netočno: C
I58	Točno: B	Točno: A	Točno: C	Netočno: B	Netočno: D	Netočno: B	Netočno: C
I59	Točno: B	Netočno: D	Netočno: A	Netočno: B	Točno: B	Točno: A	Točno: D
I60	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Netočno: A
I61	Točno: B	Točno: A	Točno: C	Netočno: B	Netočno: A	Netočno: B	Netočno: A
I62	Netočno: A	Točno: A	Netočno: A	Netočno: B	Netočno: D	Točno: A	Netočno: C
I63	Točno: B	Točno: A	Netočno: A	Netočno: B	Točno: B	Točno: A	Netočno: A
I64	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Netočno: B
I65	Točno: B	Netočno: C	Netočno: A	Točno: D	Netočno: C	Točno: A	Točno: D
I66	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
I67	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Netočno: D	Netočno: C
I68	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I69	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Točno: A	Točno: D
I70	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Točno: A	Točno: D
I71	Netočno: A	Točno: A	Netočno: D	Točno: D	Netočno: A	Netočno: D	Netočno: C
I72	Točno: B	Točno: A	Netočno: A	Netočno: B	Netočno: C	Točno: A	Točno: D
I73	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: A	Netočno: D	Netočno: A
I74	Netočno: A	Netočno: B	Netočno: A	Netočno: A	Netočno: D	Netočno: B	Netočno: B
I75	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Točno: A	Točno: D
I76	Netočno: A	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
I77	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Točno: A	Točno: D
I78	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Točno: A	Točno: D
I79	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I80	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I81	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: A	Netočno: D	Točno: D
I82	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Netočno: D	Točno: D
I83	Netočno: A	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I84	Točno: B	Netočno: C	Netočno: A	Netočno: C	Netočno: C	Netočno: B	Točno: D
I85	Točno: B	Točno: A	Netočno: A	Netočno: B	Netočno: A	Točno: A	Netočno: B

I86	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Netočno: C
I87	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Točno: D
I88	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Netočno: D	Netočno: B
I89	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Netočno: C	Točno: D
I90	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Netočno: A
I91	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I92	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: C	Točno: D
I93	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Točno: D
I94	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I95	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I96	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
I97	Točno: B	Točno: A	Netočno: A	Netočno: A	Točno: B	Točno: A	Netočno: C
I98	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Točno: A	Točno: D
I99	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Točno: A	Točno: D
I100	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Točno: A	Točno: D
I101	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Netočno: D	Točno: D
I102	Točno: B	Točno: A	Točno: C	Netočno: B	Točno: B	Točno: A	Točno: D
I103	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I104	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
I105	Točno: B	Netočno: C	Netočno: A	Netočno: B	Netočno: D	Netočno: D	Točno: D
I106	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Točno: A	Točno: D
I107	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Točno: D
I108	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Točno: A	Točno: D
I109	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Točno: A	Točno: D
I110	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
I111	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Točno: A	Točno: D
I112	Točno: B	Netočno: D	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I113	Točno: B	Netočno: B	Točno: C	Netočno: A	Netočno: D	Netočno: D	Netočno: B
I114	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I115	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Netočno: D	Točno: D
I116	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I117	Netočno: A	Točno: A	Točno: C	Netočno: B	Netočno: A	Netočno: D	Netočno: A

I118	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I119	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Netočno: D	Netočno: A
I120	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Netočno: D	Točno: D
I121	Netočno: A	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I122	Netočno: A	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Netočno: C
I123	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Netočno: D	Točno: D
I124	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Točno: D
I125	Točno: B	Netočno: C	Točno: C	Netočno: A	Točno: B	Netočno: D	Netočno: C
I126	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Točno: A	Točno: D
I127	Netočno: A	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I128	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Netočno: D	Točno: D
I129	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Točno: A	Točno: D
I130	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Točno: A	Točno: D
I131	Točno: B	Točno: A	Točno: C	Netočno: B	Točno: B	Točno: A	Točno: D
I132	Netočno: A	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
I133	Točno: B	Točno: A	Netočno: B	Točno: D	Točno: B	Točno: A	Točno: D
I134	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Točno: A	Točno: D
I135	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Točno: A	Točno: D
I136	Netočno: A	Točno: A	Netočno: D	Netočno: B	Netočno: C	Netočno: B	Točno: D
I137	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Točno: A	Točno: D
I138	Točno: B	Točno: A	Netočno: A	Netočno: B	Netočno: A	Netočno: D	Točno: D
I139	Točno: B	Točno: A	Točno: C	Netočno: B	Netočno: C	Netočno: B	Netočno: B
I140	Točno: B	Netočno: C	Točno: C	Netočno: A	Netočno: D	Netočno: B	Netočno: A
I141	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
I142	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Netočno: D	Točno: D
I143	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I144	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Netočno: D	Netočno: C
I145	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D
I146	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I147	Točno: B	Točno: A	Netočno: A	Netočno: B	Netočno: C	Točno: A	Netočno: C
I148	Netočno: A	Točno: A	Točno: C	Točno: D	Točno: B	Netočno: D	Točno: D
I149	Točno: B	Točno: A	Točno: C	Netočno: B	Netočno: C	Netočno: B	Točno: D

I150	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Točno: A	Točno: D
I151	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
I152	Točno: B	Točno: A	Netočno: D	Netočno: A	Netočno: C	Netočno: D	Točno: D
I153	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: A	Netočno: D	Točno: D
I154	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Točno: D
I155	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Točno: A	Točno: D
I156	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Netočno: D	Točno: D
I157	Netočno: D	Netočno: D	Netočno: D	Netočno: B	Netočno: A	Netočno: C	Netočno: A
I158	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
I159	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
I160	Netočno: C	Točno: A	Točno: C	Točno: D	Točno: B	Točno: A	Točno: D

Rezultati testa razumijevanja (2019.):

	Kako se osoba osjeća kada govori o svom zanimanju?	Zašto se osoba odlučila za nastavničko zanimanje?	Kakav je, prema mišljenju ove osobe, stav javnosti prema nastavničkom zanimanju?	S obzirom na opisanu situaciju, ova se osoba zalaže za to da...	Osoba smatra da se nastavnici boje...	Osoba smatra da bi bilo dobro kada bi...	Prva iskustva s vrednovanjem nastavnika pokazuju da...
i1	Netočno: A	Točno: A	Točno: C	Točno: D	Točno: B	Netočno: B	Netočno: B
i2	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
i3	Točno: B	Točno: A	Točno: C	Netočno: A	Točno: B	Netočno: D	Točno: D
i4	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Točno: A	Netočno: A
i5	Točno: B	Točno: A	Točno: C	Netočno: B	Točno: B	Netočno: D	Točno: D
i6	Točno: B	Točno: A	Točno: C	Netočno: C	Točno: B	Netočno: D	Netočno: C
i7	Netočno: C	Točno: A	Netočno: A	Netočno: B	Netočno: C	Netočno: C	Točno: D
i8	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Točno: A	Točno: D
i9	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Točno: A	Točno: D
i10	Netočno: C	Točno: A	Netočno: B	Točno: D	Točno: B	Netočno: D	Točno: D
i11	Netočno: A	Netočno: B	Točno: C	Točno: D	Točno: B	Netočno: B	Točno: D
i12	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Netočno: D	Točno: D
i13	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Točno: A	Točno: D
i14	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: C	Točno: D
i15	Netočno: A	Točno: A	Netočno: D	Netočno: B	Netočno: D	Netočno: B	Netočno: B
i16	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Točno: A	Točno: D
i17	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Netočno: C	Točno: D
i18	Točno: B	Netočno: C	Točno: C	Netočno: B	Točno: B	Netočno: B	Točno: D
i19	Točno: B	Netočno: D	Točno: C	Netočno: A	Netočno: D	Netočno: B	Netočno: C
i20	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Netočno: D	Točno: D
i21	Točno: B	Točno: A	Netočno: A	Netočno: B	Netočno: C	Netočno: D	Netočno: A
i22	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Točno: A	Točno: D
i23	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: C	Točno: A	Točno: D

i24	Netočno: A	Netočno: B	Točno: C	Točno: D	Netočno: D	Netočno: C	Točno: D
i25	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
i26	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Točno: A	Netočno: B
i27	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Točno: A	Točno: D
i28	Točno: B	Netočno: C	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
i29	Točno: B	Točno: A	Točno: C	Netočno: B	Točno: B	Netočno: D	Točno: D
i30	Točno: B	Točno: A	Netočno: D	Točno: D	Točno: B	Točno: A	Točno: D
i31	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: C	Netočno: C	Točno: D
i32	Točno: B	Točno: A	Netočno: B	Točno: D	Točno: B	Netočno: D	Točno: D
i33	Točno: B	Netočno: D	Netočno: D	Točno: D	Netočno: C	Točno: A	Točno: D
i34	Točno: B	Netočno: B	Netočno: D	Točno: D	Netočno: C	Netočno: C	Netočno: B
i35	Točno: B	Točno: A	Točno: C	Netočno: B	Netočno: D	Netočno: B	Netočno: B
i36	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Točno: A	Točno: D
i37	Točno: B	Točno: A	Točno: C	Točno: D	Točno: B	Netočno: D	Točno: D
i38	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: A	Točno: A	Točno: D
i39	Točno: B	Točno: A	Netočno: A	Točno: D	Netočno: D	Netočno: D	Točno: D
i40	Točno: B	Točno: A	Netočno: D	Točno: D	Netočno: D	Netočno: D	Netočno: A
i41	Točno: B	Točno: A	Netočno: A	Točno: D	Točno: B	Točno: A	Točno: D
i42	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: C	Netočno: D	Točno: D
i43	Točno: B	Točno: A	Netočno: B	Točno: D	Točno: B	Netočno: D	Točno: D
i44	Točno: B	Točno: A	Točno: C	Točno: D	Netočno: D	Netočno: D	Točno: D
i45	Netočno: C	Točno: A	Netočno: D	Točno: D	Netočno: D	Netočno: D	Točno: D

Prilog E

Usporedni prikaz ocjena ispitanika, broja točnih odgovora na testu razumijevanja te sigurnosti ispitanika u odabir odgovora:

	Ukupni stav	Broj točnih odgovora	Sigurnost ispitanika u odabir odgovora:
I1	3	7/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I2	4	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I3	5	1/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I4	4	6/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I5	4	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I6	4	6/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I7	3	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I8	4	7/7	Niti sam siguran/na niti nisam.
I9	3	3/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I10	4	4/7	Niti sam siguran/na niti nisam.
I11	5	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I12	5	7/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I13	4	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I14	5	6/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I15	5	6/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I16	4	6/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I17	4	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I18	4	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I19	3	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I20	3	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I21	5	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I22	4	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I23	5	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I24	5	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.

I25	3	6/7	Sasvim sam siguran/na u odabir odgovora.
I26	4	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I27	5	3/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I28	4	6/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I29	4	4/7	Niti sam siguran/na niti nisam.
I30	4	5/7	Pretpostavljam da sam odabrao/la većinu netočnih odgovora.
I31	4	7/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I32	4	3/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I33	4	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I34	5	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I35	4	4/7	Niti sam siguran/na niti nisam.
I36	4	3/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I37	5	7/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I38	3	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I39	4	3/7	Niti sam siguran/na niti nisam.
I40	3	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I41	4	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I42	4	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I43	4	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I44	3	3/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I45	4	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I46	4	3/7	Niti sam siguran/na niti nisam.
I47	5	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I48	5	7/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I49	5	5/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I50	4	3/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I51	4	7/7	Pretpostavljam da sam odabrao/la većinu netočnih odgovora.
I52	4	4/7	Niti sam siguran/na niti nisam.
I53	4	5/7	Niti sam siguran/na niti nisam.
I54	4	3/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I55	4	3/7	Niti sam siguran/na niti nisam.
I56	4	5/7	Niti sam siguran/na niti nisam.

I57	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I58	4	3/7	Niti sam siguran/na niti nisam.
I59	5	4/7	Nimalo nisam siguran/na u odabir odgovora.
I60	5	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I61	4	3/7	Niti sam siguran/na niti nisam.
I62	5	2/7	Niti sam siguran/na niti nisam.
I63	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I64	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I65	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I66	5	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I67	5	3/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I68	5	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I69	4	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I70	4	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I71	4	2/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I72	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I73	4	3/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I74	4	0/7	Sasvim sam siguran/na u odabir odgovora.
I75	5	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I76	5	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I77	4	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I78	5	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I79	4	7/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I80	4	7/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I81	4	5/7	Niti sam siguran/na niti nisam.
I82	3	4/7	Niti sam siguran/na niti nisam.
I83	3	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I84	2	2/7	Sasvim sam siguran/na u odabir odgovora.
I85	3	3/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I86	5	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I87	5	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I88	3	3/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.

I89	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I90	2	4/7	Pretpostavljam da sam odabrao/la većinu netočnih odgovora.
I91	4	7/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I92	1	5/7	Pretpostavljam da sam odabrao/la većinu netočnih odgovora.
I93	5	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I94	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I95	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I96	4	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I97	5	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I98	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I99	4	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I100	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I101	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I102	5	6/7	Niti sam siguran/na niti nisam.
I103	5	7/7	Niti sam siguran/na niti nisam.
I104	4	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I105	3	2/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I106	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I107	3	5/7	Niti sam siguran/na niti nisam.
I108	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I109	3	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I110	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I111	3	5/7	Niti sam siguran/na niti nisam.
I112	2	6/7	Niti sam siguran/na niti nisam.
I113	4	2/7	Niti sam siguran/na niti nisam.
I114	2	7/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I115	4	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I116	3	7/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I117	4	2/7	Pretpostavljam da sam odabrao/la većinu netočnih odgovora.
I118	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I119	3	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I120	3	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.

I121	4	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I122	2	3/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I123	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I124	4	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I125	5	3/7	Nimalo nisam siguran/na u odabir odgovora.
I126	5	6/7	Sasvim sam siguran/na u odabir odgovora.
I127	3	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I128	4	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I129	5	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I130	4	5/7	Niti sam siguran/na niti nisam.
I131	4	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I132	5	4/7	Niti sam siguran/na niti nisam.
I133	4	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I134	3	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I135	4	5/7	Niti sam siguran/na niti nisam.
I136	3	2/7	Niti sam siguran/na niti nisam.
I137	3	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I138	5	3/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I139	3	3/7	Niti sam siguran/na niti nisam.
I140	4	2/7	Pretpostavljam da sam odabrao/la većinu netočnih odgovora.
I141	5	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I142	3	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I143	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I144	5	3/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I145	4	7/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I146	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I147	3	3/7	Niti sam siguran/na niti nisam.
I148	4	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I149	3	4/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I150	4	5/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I151	3	6/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.
I152	3	3/7	Pretpostavljam da sam odabrao/la većinu tačnih odgovora.

I153	5	5/7	Niti sam siguran/na niti nisam.
I154	3	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I155	5	6/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I156	5	4/7	Niti sam siguran/na niti nisam.
I157	3	0/7	Nimalo nisam siguran/na u odabir odgovora.
I158	2	4/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.
I159	5	4/7	Niti sam siguran/na niti nisam.
I160	2	6/7	Pretpostavljam da sam odabrao/la većinu točnih odgovora.

Životopis

Sandra Ljubas rođena je 1992. godine u Pforzheimu, SR Njemačka. Nakon završetka osnovne škole i IV. gimnazije u Zagrebu upisuje dvopredmetni preddiplomski studij Njemačkog jezika i književnosti te Švedskog jezika i kulture na Filozofskom fakultetu Sveučilišta u Zagrebu, a potom i prevoditeljske smjerove diplomskih studija na Odsjeku za germanistiku i Katedri za skandinavistiku istog fakulteta. Obrativši oba prevoditeljska diplomatska rada uspješno 2017. godine stječe naziv *mag. philol. germ. et suet.*, a odlikovana je i Nagradom za izvrsnost u studiju. Već 2018. upisuje Poslijediplomski doktorski studij Humanističkih znanosti na Sveučilištu u Zadru, smjer Filologija – jezikoslovlje. Suautorica je *Švedsko-hrvatskog rječnika leksičkih inovacija*, bila je stručna suradnica u izradi *Hrvatsko-švedskog razgovornog priručnika* i objavila je nekoliko znanstvenih radova iz područja lingvistike i znanosti o prevođenju. Glavne istraživačke interese čine joj strojno prevođenje i vrednovanje strojnog prevođenja, leksikologija, kao i razne traduktološke teme. Radi u Zagrebu kao slobodna prevoditeljica, prvenstveno se baveći književnim prevođenjem, a dosad je objavljeno dvadesetak njezinih romana u prijevodu. Prevodi s njemačkog, švedskog, norveškog i danskog jezika. Usavršavala se na brojnim jezičnim seminarima i radionicama prevođenja u Austriji i Njemačkoj te Švedskoj i Danskoj. Aktivna je članica Društva hrvatskih književnih prevodilaca i Hrvatske zajednice samostalnih umjetnika. Kao stipendistica programa Artist-in-Residence Litteratur Västra Götaland 2020. provodi mjesec dana u istraživačko-prevoditeljskoj rezidenciji u Jonseredu, u Švedskoj. Prosinac 2021. i siječanj 2022. provodi u rezidenciji za pisce i prevoditelje Baltic Centre for Writers and Translators na otoku Gotlandu, u Švedskoj.