

## Gradnja in analiza petjezičnega korpusa podnaslovov govorov TED\*

Miha Helbl,<sup>†</sup> Žiga Domevšček<sup>‡</sup>

<sup>†</sup> Oddelek za prevajalstvo, Filozofska fakulteta, Univerza v Ljubljani

Aškerčeva 2, 1000 Ljubljana

<sup>†</sup> miha.helbl@gmail.com, <sup>‡</sup> ziga.domevscek@gmail.com

### Povzetek

Namen članka je preučevanje funkcionalnosti prevoda slovenskih prostovoljnih prevajalcev za nepridobitno organizacijo TED, ki niso nujno strokovno izobraženi prevajalci, temveč zgolj ljubitelji ali poznavalci teme govora. Analizirali smo odstopanja med angleškimi izvirmiki in njihovimi slovenskimi prevodi. Analiza obsega pregled prevajalskih strategij posploševanja, konkretizacije, zamenjave samostalnika z zaimkom ter izpustov znotraj slovenskih in angleških segmentov izbranih reprezentativnih govorov. Za namen raziskave so v petjezični korpus vključene izvorne angleške transkripcije ter slovenski, nemški, italijanski in francoski podnaslovi analiziranih govorov. Pričakujemo, da bodo prevajalci za doseganje ustreznosti prevodnih rešitev pogosto nezavedno uporabljali vrsto različnih prevajalskih strategij.

### Compilation and Analysis of a Multilingual Corpus of TED Subtitles

The purpose of this article was to analyse the functionality of the Slovenian translations made by volunteers for the non-profit organization TED. These volunteers are not necessarily educated as translators, but are often fans or experts on the talk topic. We analysed the discrepancies between the English originals and their Slovenian translations. The focus was on the following translation strategies: generalization, specification, substitution of the noun with a pronoun, and omissions. The multi lingual corpus, created for the purpose of this research, consists of the original English transcripts and the corresponding Slovenian, French, Italian and German subtitles. We expect that the translators, in order to achieve the suitable translational solution, will often unknowingly use a variety of translational strategies.

### 1 Uvod

Projekt gradnje in analize petjezičnega korpusa podnaslovov govorov TED je bil zasnovan z namenom raziskave prevedene vsebine s strani prostovoljnih prevajalcev, ki opravljajo delo za nepridobitno organizacijo TED. Želeli smo preučiti funkcionalnost prevodov slovenskih prostovoljnih prevajalcev, ki niso zagotovljeni šolani prevajalci ali jezikoslovci, temveč so najpogosteje zgolj ljubitelji ali poznavalci teme določenega govora. V ta namen smo zbrali podnaslove 316 govorov v petih različnih jezikih. Glavni kriterij izbora je bil, da so podnaslovi prevedeni v slovenski jezik iz angleškega izvirnika. Pri analizi smo uporabili samo slovenske in angleške podnaslove. Želeli smo prikazati razhajanja med izvornimi angleškimi transkripcijami govorov TED in njihovimi slovenskimi prevodi na podlagi odstopanja pri izpuščanju in zgoščevanju. Z analizo teh odstopanj smo želeli izpostaviti najpogostejše sporne segmente pri prevajanju in pretvorbi transkripcij v podnaslove.

Nepridobitna organizacija TED je bila ustanovljena leta 1984 s ciljem, da bi širila ideje na področjih tehnologije, zabave in oblikovanja (Technology, Entertainment, Design). Danes njene konference

pokrivajo skoraj vsa področja od znanosti do gospodarstva in do globalnih vprašanj in so od junija 2006 na voljo za sprotno gledanje prek interneta. Spletni arhiv TED omogoča ogled govorov s podnaslovi, ki jih prispevajo prostovoljni prevajalci. Skupno si je od januarja 2016 na spletišču TED možno ogledati govore s podnaslovi v 110 jezikih, ki jih je pomagalo prevesti 16.114 prostovoljnih prevajalcev. Tako imenovani Projekt odprtega prevajanja (Open Translation Project) je TED vzpostavil na pobudo obiskovalcev spletišča, ki so želeli TED govore deliti s svojimi prijatelji in družinskimi člani in so organizatorje vprašali, če lahko oni sami prevedejo te govore in jih opremijo s podnaslovi. Organizacija TED je v ta namen vzpostavila sistem, ki je omogočil prostovoljcem prevajanje njihovih najljubših govorov v poljuben jezik. V času opravljanja raziskave je bilo del prevajalske ekipe tudi 106 prevajalcev za slovenščino, ki so prevedli več kot 300 govorov.<sup>1</sup>

Na svojem spletišču so organizatorji objavili tudi smernice za vse prevajalce, da bi tako uskladili podnaslove v vseh jezikih. Prva in hkrati osrednja točka smernic je sodelovanje najmanj dveh prostovoljcev pri prevajalskem projektu – prevajalca in lektorja. Druga zajema problematiko podnaslavljanja in prostovoljcem predstavi standarde glede dolžine podnaslovov. Tretja

\* Projekt gradnje korpusa je potekal v okviru magistrskega modula Korpusi in baze podatkov na Oddelku za prevajalstvo FF UL (štud. l. 2015/2016) pod mentorskim vodstvom doc. dr. Darje Fišer. Besedilo je nastalo pri predmetu Slovensko strokovno besedilo pri prof. dr. Vojku Gorjancu.

<sup>1</sup> TED. <https://www.ted.com/people/translators?sort=translations&country=slovenia> (Dostop 31. 1. 2016)

pa zajema slog TED, ki je osrednji slogovni del smernic in med drugim spodbuja uporabo neformalnega, sproščene in vsebinsko ter jezikovno natančnega sloga prevajanja. Prevajalci imajo poleg smernic v besedilu na voljo tudi vodič v video formatu.

Tak način prevajanja spada med tako imenovano množičeno prevajanje (crowdsourced translation). Množičeno prevajanje izkoristi interes, ki združuje člane prevajalskih projektov. Ti člani sodelujejo pri prevajanju, lektoriranju, zbiranju idej, virov in gradnji pomnilnikov prevodov. Množičenje takšnim skupinam odpira možnosti za nudenje nadaljnjih prevajalskih storitev, ki so lahko ali plačana ali pa delujejo na prostovoljski ravni.

V članku je predstavljen pregled področja in raziskave, ki so proučevale podnaslovno prevajanje. V teoretičnemu delu nato sledi zasnova projekta, prikaz najpogostejšega besedja in analiza prevodov.

## 2 Pregled področja in sorodne raziskave

Delia Chiaro (2013) označi avdiovizualno prevajanje kot prenos verbalnih prvin, ki jih nosijo avdiovizualna dela in produkti (filmi, opere, televizijske oddaje, muzikali, računalniške igre itn.) iz enega jezika v drugega. Razlikuje med dvema glavnima načinoma avdiovizualnega prevajanja – sinhronizacijo in podnaslavljanjem. Podnaslovi so skrajšani zapisani prevodi tistega, kar je moč slišati na ekranu. S takim izhodiščem Mona Baker (1992) definira sedem prevajalskih strategij – posploševanje, nevtralizacija, kulturna substitucija, kalkiran prevod z razlago, parafraziranje, izpust in prevod s sliko – od katerih sta za našo analizo pomembna predvsem posploševanje in izpust. Mona Baker označuje posploševanje kot eno od najsplošnejših strategij pri prevajanju, kjer ni ustreznice v ciljnem jeziku. Ta strategija je znotraj večine jezikov enako učinkovita, saj hierarhična struktura semantičnih polj ni vezana na določen jezik (Baker, 1992: 26). Če pomen določenega izraza ni ključen za razvoj nadaljnjega besedila in bi njegova razlaga zgolj zmotila bralca, se prevajalci pogosto lahko odločijo za izpust prevoda besede ali izraza (Baker, 1992: 40).

Na slovenskem področju je Urša Vogrinc Javoršek (2007) izvedla raziskavo, kjer je proučevala zgolj metabesedilne elemente v igrani ameriški nanizanki *Midve z mamami* (Gilmore girls). Metabesedilni elementi so tisti deli, ki ne prispevajo k vsebini besedila, temveč pomagajo gledalcu vsebino bolje in lažje organizirati ter razložiti (Vogrinc Javoršek po Halliday, 1985: 35). Izsledki raziskave so pokazali, da so podnaslovi prevodi igrane nanizanke manj kohezivni kot izvirnik. To drži, če upoštevamo zgolj verbalni vidik, a podnaslovi nastanejo na podlagi dokončanega avdiovizualnega medija, kjer je potrebno upoštevati tudi zvok in sliko. Prevajalec tako upošteva dejstvo, da informacije ne prihajajo zgolj skozi en kanal. Dopusča, da se določeni elementi, ki se pojavijo v drugih kanalih,

ne obdržijo v podnaslovu (Vogrinc Javoršek, 2007). Irena Kovačič je več kot desetletje pred tem (1995) izpostavila, da današnje norme v podnaslovih zelo očitno odsevajo odnos do jezika in še posebno do registra, ki ignorira najdbe v moderni lingvistiki, zlasti na področju razlik med pisnim in govornim jezikom ali funkcionalnimi vrednostmi različnih jezikovnih struktur. Razlogi za to so očitni: podnaslovi so prikazani v pisni obliki, ki je že po naravi bolj organizirana in bolj zgoščena kot govor, tako je idealna za podnaslavljanje, kjer je redukcija oblik ena osnovnih zahtev. Pisni jezik je relativno standardiziran, medtem ko govorni jezik ni. Prevajalci se morajo ne le odločiti, kako prevesti besedilo kot celoto in kako prenesti elemente v ciljni jezik, ampak v prvi vrsti se morajo tudi odločiti, kaj sploh prevesti in kaj izpustiti (Kovačič, 1996). Te odločitve so tesno povezane z dvema določenima lastnostma podnaslavljanja: a) potreba po zgoščevanju sporočila zaradi omejenega prostora, b) prenos iz govornega v pisni diskurz. Kovačič po Hallidayju definira tri funkcije zgoščevanja – (med)osebno, besedilno in ideacijsko. Ideacijska funkcija je od naštetih najbolj odvisna od jezika, medosebna je pogosto replicirana v neverbalni in nejezikovni interakciji. Besedilna funkcija pa postane pri podnaslovih drugotnega pomena, ker slika poskrbi za kontinuiteto in kohezijo. Prevajalska strategija je odvisna od vrste programa in pričakovanega ciljnega občinstva. Za ciljno občinstvo govorov TED se s strani organizacije predvideva, da imajo sposobnost hitrejšega branja, širše besedišče, je že seznanjeno s tematiko ali pa pozna ozadje govorov. Zato so podnaslovi lahko gostejši, vsebujejo več dobesednih izrazov, hkrati pa morajo izraziti ne zgolj zgodbo ampak tudi osebnosti govorcev ter psihološke odnose in pozicije moči med osebami (Kovačič, 1996).

Pri podnaslovnem prevajanju prihaja do kompromisa med normami pisanega in govornega jezika pri čemer se prevod in priredba zlijeta in potekata istočasno. Pri tem naj bi se prevajalci trudili, da uporabijo polno strukturo, če je le mogoče (Halliday, 1994). Organizacija TED se v svojih smernicah za prevajalce manj ukvarja s strukturo kot s pomenom, saj med drugim narekuje da naj prevajalci kljub časovnim in prostorskim omejitvam predvsem natančno posredujejo pomen govornega besedila. Tu moramo opozoriti na to, da podnaslovi govorov TED tako po času trajanja in dolžini posamezne vrstice odstopajo od ustaljenih norm, ki se uporabljajo pri televizijskih in kinematografskih podnaslovih pri nas. Prav tako odstopajo po najvišjem dovoljenem številu znakov na sekundo. Tako TED v svojih smernicah prevajalcem<sup>2</sup> narekuje, da je v posamezni vrstici dovoljenih največ 42 znakov, pri čemer prevajalec ne sme vstaviti več kot dveh vrstic. Najvišja dovoljena bralna hitrost prikazovanja podnaslovov je po normah njihovih smernic 21 znakov na sekundo. Slovenski standardi za

<sup>2</sup> TED. <https://www.ted.com/participate/translate/guidelines>. (Dostop 31. 1. 2016)

podnaslavljanje, ki jih je tudi moč zaslediti pri nekaterih televizijskih hišah in prevajalskih agencijah (na primer prevajalski studio Milenko Babič, s. p.), v povprečju narekujejo največ 36 znakov na posamezno vrstico in bralno hitrost 12 do 15 znakov na sekundo. Ti standardi se lahko razlikujejo med naročniki.<sup>3</sup>

Da pa se prevajalci lahko držijo zgoraj omenjenih prostorsko-časovnih omejitev, morajo prevodno besedilo ustrezno prilagoditi. Pri tem si pomagajo z zgoraj omenjenimi prevajalskimi strategijami. Mi smo se pri naši analizi osredotočili na posploševanje, konkretizacijo, zamenjave samostalnika z zaimkom ter izpuste, ki smo jih razdelili na tri sklope: vsebinske izpuste, izpuste diskurzivnih označevalcev in izpuste ponavljanj. Odločitev za opazovanje diskurzivnih označevalcev utemeljujemo z njihovo naravo, saj so to jezikovna sredstva, ki imajo v prvi vrsti pragmatično vlogo (niso tako pomembni za posredovanje informacij ter vsebine, ampak razvijajo medosebni odnos sogovornikov pri organizaciji poteka diskurza (Verdonik, 2006), označujejo prehode med tematskimi sklopi (prav, okej, no, dobro, v redu ...), prehod v zaključek govora (dobro, okej, prav, torej), popraviljanje ali druge spremembe v strukturi izjave ... Gre torej za jezikovne elemente, za katere lahko glede na napotke pri TED-u, pa tudi zaradi omejitve pri prostoru in povezave med sliko in podnaslovom, pričakujemo, da jih bodo prevajalci v večji meri izpuščali.

### 3 Zasnova projekta

Projektno delo je obsegalo gradnjo korpusa, pridobivanje metapodatkov in dve jezikoslovno-prevodni analizi. Za osnovo korpusa smo zbrali vse slovenske podnaslove na spletišču TED, ki jih je bilo v času pisanja članka 316. Prenesli smo slovenske, angleške, nemške, francoske in italijanske podnaslove za vsakega od 316 govorov. Vse govore smo zbrali v arhiv in jih razporedili glede na leto nastanka. Vsi podnaslovi so vsebovali tudi časovne kode, ki nas pri gradnji in analizi korpusa niso zanimale, zato smo iz njih ustvarili ločen arhiv.

Hkrati smo na spletišču TED za vseh pet jezikov izpisali tudi metapodatke, ki bi jih vključili v poseben korpus. Med te podatke smo vključili naslov govora, ime govorca, prevajalca, lektorja, opis govora, leto in mesec konference ter ime konference na kateri je bil govor predstavljen. Metapodatki bodo služili kot dodaten referenčni podatkovni material pri uporabi korpusa.

Ustvariti smo želeli paralelni korpus, ki vsebuje poravnana besedila v več jezikih. Za poravnavo podnaslovov smo uporabili orodje LF Aligner. V orodje smo za vsak posamezen govor naložili podnaslove v vseh petih jezikih v obliki .TXT. Program nam je podnaslove poravnal in ustvaril en sam Excelov dokument z vsemi petimi podnaslovi, poravnanimi horizontalno, ter tudi pomnilnik prevodov v datotečni

obliki .TMX. Postopek smo ponovili za vseh 316 govorov. Pomnilnike prevodov smo naložili na odprto korpusno spletišče SketchEngine, kjer smo paralelni korpus ustvarili za vseh pet jezikov. Zaradi omejenosti dostopa do korpusa v času raziskave smo bili s paralelnim korpusom omejeni zgolj na milijon besed. V tistem času še nismo vedeli, da bi lahko prosili za povečanje pomnilnika in bi tako lahko naložili celotno zbrano zbirko besedil – od takrat je bil pomnilnik povečan na štiri milijone besed in je že pripravljen na razširitev raziskave. Korpus smo zapolnili s 131 pomnilniki slovenskih in angleških prevodov. Ker so govori nastali v različnih letih – od leta 2004 do 2015 – smo za korpus naredili selektiven izbor govorov glede na nastanek. Kronološko smo tako izbrali prvih 30 govorov v arhivu, preskočili naslednjih 60, ponovno vključili naslednjih 30 govorov in postopek ponavljali dokler nismo korpusa napolnili do omejene količine podatkov, pri čemer smo pazili tudi na zastopanost prevodov iz različnih časovnih obdobj.

Nastali korpus smo najprej analizirali glede na obseg celotnega korpusa, na število pojavnice za posamezni jezik in opravili pregled, v okviru katerega smo opazovali najpogostejše polnopomenske besede v slovenščini in angleščini, v nadaljevanju pa smo se osredotočili na analizo prevodnih rešitev. Pri tem pa je bil naš glavni namen pregled odstopanj med izvornikom in prevodom zaradi izpustov in posploševanja besedila. Izpuste smo pri analizi razdelili na tri sklope: izpuste diskurzivnih označevalcev, izpuste ponavljanja in vsebinske izpuste. Končni cilj je bil določiti funkcionalnost prevodov s strani prostovoljnih prevajalcev, ki morebiti niso vedno izobraženi prevajalci. Pri analizi smo se omejili na govore, kjer so bila največja odstopanja pri številu segmentov. Opazili smo namreč, da je prišlo do bistvenih razlik v številu segmentov le pri omejenem številu govorov. Kot ločen segment smo opredelili en podnaslov, ki je lahko tako enovrstičen kot dvovrstičen.

Že pred začetkom analize smo pričakovali, da bo zaradi prostovoljne narave prevajalskega projekta pri organizaciji TED kljub natančnim navodilom s strani organizacije prihajalo do odstopanj pri številu segmentov slovenskih podnaslovov. Predpostavljali smo tudi, da bo v primerjavi z angleškim izvornikom v slovenskih podnaslovih pogosteje prihajalo do izpuščanja besedila kot dopolnjevanja, saj je besedilo kognitivno lažje izpuščati kot pa dodajati.

### 4 Analiza

Analiza korpusa je sestavljena iz dveh delov. Najprej smo opravili krajši pregled najpogostejšega besedišča v zgrajenem vzporednem korpusu govorov TED za slovenski in angleški jezik v konkordančniku SketchEngine. Med analizo korpusa smo morali rešiti dilemo, kako ravnati, saj slovenski nabor besedil ni bil jezikovno označen (med opravljanjem raziskave nismo bili seznanjeni s slovenskim označevalnikom). Zaradi

<sup>3</sup> Lektorsko društvo Slovenije. <http://www.lektorsko-drustvo.si/predavanja/na-istem-bregu>. (Dostop 31. 1. 2016)

tega ni bilo možno izvesti popolnoma reprezentativnega prikaza slovenskega korpusa. Slovenski prikaz najpogostejšega besedišča namreč ne odseva najpogostejših pojavnic, temveč najpogostejše različnice. Tako smo poiskali deset najpogostejših slovenskih različnic (samostalnike, glagole, pridevnike in prislove), ki so se pojavile na frekvenčnem seznamu in smo opravili prevodno primerjavo z angleškimi pojavnicami. Prikaz angleškega korpusa pa je bolj reprezentativen, saj je za razliko od slovenskega korpusa označen in smo zato zbrali deset najpogostejših pojavnic, ki smo jih tudi označili glede na besedno vrsto.

Sledila je analiza prevodov, kjer smo opazovali najpogostejše tipe uporabljenih prevajalskih strategij, hkrati pa smo v sprotni diskusiji podali lastna domnevanja, zakaj je do teh posegov v besedilo med prevajanjem prišlo.

#### 4.1 Prikaz najpogostejšega besedišča

V celotnem petjezičnem korpusu je 131 govorov za vsak posamezen jezik. Skupno teh 131 govorov v petih jezikih zajema 968.559 različnic, kar je zgolj tretjina celotnega korpusnega arhiva 316 govorov. Od tega smo se za analizo osredotočili na slovenski in angleški del. Skupno število slovenskih besed znaša 181.536 različnic oziroma 222.276 pojavnic, medtem ko število angleških besed znaša 226.935 različnic oziroma 270.741 pojavnic, skupna velikost slovensko-angleškega korpusa je potemtakem 408.471 besed. Že po končnem številu besed za posamezen jezik je razvidno, da je bilo v slovenskih podnaslovih veliko izpuščanja, kar zgolj potrjuje obe naši začetni hipotezi.

	Pojavnica	Besedna oblika	Število
1	so	RB	1629
2	people	NNS	866
3	do	VV	865
4	now	RB	703
5	just	RB	682
6	very	RB	634
7	really	RB	516
8	going	VVG	496
9	then	RB	465
10	know	VVP	463

Tabela 1: Najpogostejše angleške pojavnice.

	Različnica	Število
1	ljudje	816
2	stvar	505
3	smeh	396
4	videti	323

5	način	254
6	aplavz	235
7	leta	201
8	hvala	193
9	primer	136
10	svet	122

Tabela 2: Najpogostejše slovenske različnice.

Najpogostejše pojavnice v angleškem in različnicah slovenskem jeziku so prikazane v tabelah. V tabeli z angleškim naborom (gl. Tabela 1) so v drugem stolpcu zapisane pojavnice, ki so razvrščene po pogostosti, tretji stolpec označuje besedno obliko pojavnic v konkordančniku SketchEngine<sup>4</sup>, četrti stolpec pa prikazuje absolutno frekvenco pojavnice. Najvišjo frekvenco pojavitev so pri tem imeli prislovi.

Seznam najpogostejših polnopomenskih angleških pojavnic smo poiskali s pomočjo funkcije *word list* znotraj konkordančnika, kjer smo lahko besede razdelili glede na oznako. Iskanje najpogostejših polnopomenskih slovenskih različnic pa je tako rekoč potekalo ročno, saj kot že omenjeno, slovenski korpus v času opravljanja raziskave še ni bil označen. Nabor najpogostejših različnic je prikazan v Tabeli 2.

Po primerjavi zbranih najpogostejših pojavnic in različnic ter prevodni primerjavi besedil znotraj korpusa smo ugotovili, da odstopanja pri številu najpogostejših besed niso tako velika. Glavne polnopomenske besede imajo odstopanja do približno 100 besed, kar je deloma tudi posledica tega, da nismo zbrali celotnega nabora besed.

#### 4.2 Analiza prevodov podnaslovov

Analizo prevodov smo razdelili na šest večjih sklopov. Pri vsakem sklopu bomo podali po pet primerov za določen pojav, hkrati pa bomo ob tem ugotavljali, zakaj je do teh odstopanj prišlo. Sklopi so razporejeni po pogostosti pojavitev v podnaslovih od najpogostejših do najredkejših.

Iz dvajsetih govorov je bilo skupno izpisanih 202 primera odstopanja. Najpogostejši so bili izpusti diskurzivnih označevalcev s 46 odstotki, sledili so jim vsebinski izpusti s slabimi 20 odstotki, posploševanje s 14 odstotki, ostale strategije pa so zamenjave samostalnika z zaimkom z 11 odstotki, ponavljanje s 5 odstotki in konkretizacija s 4 odstotki. Skupno je bilo izpustov 71 odstotkov, ostalih strategij pa 29 odstotkov.

Največ odstopanja pri slovenskih in angleških podnaslovih je bilo pri izpuščanju diskurzivnih označevalcev v slovenskih podnaslovih. Do tega je prišlo zaradi razlikovanj v primarni funkciji besedila, saj je angleški del bolj transkripcija govora, slovenski pa je dejanski podnaslov. Poleg tega so diskurzivni označevalci značilni za govor, saj povezujejo

<sup>4</sup> SketchEngine. <https://www.sketchengine.co.uk/penn-treebank-tagset/>. (Dostop 31. 1. 2016)

posamezne segmente le-tega, imajo zelo nizko informacijsko vrednost in so pri krašjanju besedila za podnaslov prvi, ki odpadejo. Prevod označevalcev ni prioriteten, kar je lepo razvidno iz prvega primera, kjer je *well*, slovensko *torej* izpuščen, saj gledalcu ne poda nove informacije (gl. Tabela 3).

Angleščina	Slovenščina
Well it's very simple	To je enostavno.
And then finally – you know, /.../	Na koncu --najboljša možnost-- /.../
Now, when psychologists show you bars, /.../	Ko vam psihologi kažejo stolpce, /.../
So thank you very much.	Najlepša hvala.
»Oh, I created this great product, /.../«	»Ustvaril sem odličen produkt, /.../«

Tabela 3: Primeri izpuščanja diskurzivnih označevalcev.

Drugi najpogostejši način krašjanja besedila so bili vsebinski izpusti, kjer je gledalcu zamolčan del informacij, ki lahko vplivajo na razumevanje podnaslova. To je dobro razvidno pri prvem primeru, kjer je pridevnik, ki določa tip očal, zamolčan. Morda je bil to zgolj lapsus ali pa so *varnostna* očala bila predolga za podnaslov (gl. Tabela 4).

Angleščina	Slovenščina
/.../ a pair of safety glasses, /.../	/.../ ker lahko s parom očal, /.../
/.../ extra-virgin olive oils /.../	/.../ vrst ekstra deviškega olja /.../
My goal is to better understand these violent actors /.../	Ta nasilna gibanja želim bolje razumeti /.../
/.../ that they can actually have experiences in their heads /.../	/.../ da lahko dejansko doživlja dogodke /.../
/.../ from your life.	/

Tabela 4: Primeri vsebinskega izpuščanja.

Posploševanje je tretji najpogostejši pojav krašjanja besedila. Do njega pride takrat, ko je v izvorniku opisana točno določena stvar, ki morda naši kulturi ni najbližja ali pa bi bil njen prevod preprosto predolg. To lepo ponazarja prvi primer, kjer v izvorniku rečejo *pop quiz*, kar pomeni nenapovedan preizkus znanja, slovenski prevajalec pa se odloči zgolj za *preizkus znanja*, saj bi bil točen prevod predolg (gl. Tabela 5).

Angleščina	Slovenščina
You failed the pop quiz, and you're hardly five minutes into the lecture.	To je enostavno. Po petih minutah predstavitve ste že padli na preizkusu znanja.
(Laughter) Because she was the main reason we were leaving the country.	Na koncu --najboljša možnost-- ker je bila ona glavni razlog za naš odhod.
and it's not because they whipped some up, tried it and went, "Yuck."	Ko vam psihologi kažejo stolpce, Ne zato, ker bi ga kdo naredil, poizkusil in se zgrozil.
so that they can feel better about the worlds in which they find themselves.	/.../ zato, da se lahko bolje počuti v danih okoliščinah.
Like Sir Thomas, you have this machine.	Kot g. Thomas imate tudi v to sposobnost.

Tabela 5: Primeri posploševanja.

Pri zamenjavah samostalnika z zaimkom, ki je po pogostosti na četrtem mestu, gre za podajanje že znanih informacij v krajši obliki. Pri prvem primeru že vemo, da je govora o profesorjih, zato tega ni potrebno ponovno napisati. Podnaslov razumemo, hkrati pa ga tudi preberemo hitreje (gl. Tabela 6).

Angleščina	Slovenščina
There's something curious about professors in my experience -- /.../	Nekaj zelo zanimivega je na njih po mojih izkušnjah -- /.../
But if I do this with amnesiac patients, /.../	Na koncu --najboljša možnost-- Če pa to naredim pri naših bolnikih, /.../
He has been working on the topic for 20 years.	Ko vam psihologi kažejo stolpce, Na tem je delal že 20 let.
We need to know what makes These organisations tick.	Vedeti moramo, kaj jih vodi.
On the morning of the hemorrhage, I could not walk, talk, read, write or recall any of my life.	To jutro nisem mogla hoditi, govoriti, brati, pisati ali se spomniti česar koli v svojem življenju.

Tabela 6: Primeri zamenjave samostalnika z zaimkom.

Ponavljjanje, peti najpogostejši način krašjanja, se pri podnaslovih zaradi časovno-prostorskih omejitev večinoma izpusti, saj je poglobljena funkcija podnaslovov podajanje informacij in ne ponazarjanje

stila govora govorca. Krajše je boljše. Pri prvem primeru vidimo, da se *miselni procesi* pri drugi omembi izpustijo (gl. Tabela 7).

Angleščina	Slovenščina
A system of cognitive processes, largely non-conscious cognitive processes, that help them change	Sistem miselnih procesov, večinoma nezavednih, /.../
Third: make somebody else really, really rich.	Tretje: Nekoga drugega naredite zelo bogatega.
/.../ where a tiny, tiny fraction of the world /.../	/.../ kjer se lahko le zelo majhen del sveta /.../
WK: In Malawi, Kasungu. In Kasungu.	WK: V Malaviju v mestu Kasungu.
You see the thing is, the thing is /.../	Vidite, stvar je v tem /.../

Tabela 7: Primer izpuščanja ponavljanj.

Pri konkretizaciji, ki je na zadnjem mestu po pogostosti, gre predvsem za jasnejše sporočanje izvirnika. Najbolj ilustrativen je četrti primer, kjer je slovenski prevajalec zamenjal izvorni angleški *he* s točno določenim samostalnikom, ki podrobneje opisuje njega, torej *kolega* (gl. Tabela 8).

Angleščina	Slovenščina
Here's two different futures that I invite you to contemplate.	To je enostavno. Razmislite o dveh različnih življenjskih poteh, /.../
from the one they like the most, to the one they like the least.	Na koncu --najboljša možnost-- po vrsti, glede na to, koliko so mu všeč.
/.../ from the one you like the most to the one you like the least."	Ko vam psihologi kažejo stolpce, slike glede na to, koliko so vam všeč."
So he recognizes that I need help and he gets me help.	Najlepša hvala. Kolega dojame, da potrebujem pomoč in mi jo pošlje.

Tabela 8: Primeri konkretizacije.

## 5 Zaključek

Petjezični korpus sestavljajo prevodi v slovenščini, nemščini, francoščini, italijanščini ter angleško izvorno besedilo. Med analizo prevodov smo prišli do zaključkov, da se prostovoljni prevajalci za TED med

prevajanjem sicer pogosto poslužijo strategij posploševanja, konkretizacije, zamenjave samostalnika z zaimkom ter izpustov, toda ne v tolikšni meri, kot smo prvotno pričakovali. K temu so v večji meri pripomogla prevajalcem ustrezno predstavljena navodila na spletišču TED ter dobro zamišljen sistem sodelovanja več ljudi pri posameznih prevajalskih projektih – prevajalec, lektor in tudi koordinator posamezne prevajalske skupnosti.

Pri analizi prevodov je bila najpogostejša prevajalska strategija izpust diskurzivnega označevalca, na drugem mestu pa je bil vsebinski izpust. Sledili so jim posploševanje, zamenjava samostalnika z zaimkom, izpust ponavljanja in konkretizacija. Iz tega je razvidno, da se prostovoljni prevajalci, tudi tisti, ki niso strokovno izobraženi prevajalci, poslužujejo velikega števila ustreznih prevajalskih strategij za podnaslavljanje. Ni pa jasno ali se teh strategij poslužujejo zavedno ali nezavedno, kar odpira možnosti za nadaljnje raziskave v tej smeri.

Izdelava tovrstnih specializiranih korpusov vsekakor odpira mnogo možnosti za prevajalsko stroko, saj bi lahko služili kot jezikoslovni referenčni vir za prevajalce na posameznih področjih. Že ta korpus odpira možnosti za boljše usklajevanje prevodov celotne ekipe prevajalcev za TED, saj jim daje vpogled v bazo podatkov, ki je usmerjena zgolj za njihovo področje. Vsekakor pa obstaja priložnost, da se korpus dopolni s preostalimi govori in ostalimi prevodi govorov, saj je bila v korpus bila vključena le dobra tretjina vseh zbranih govorov za angleški in slovenski jezik.

## 6 Literatura

- Mona Baker. 1992. *In Other Words. A Course in Translation*. London and New York: Routledge.
- Delia Carmela Chiaro 2012. *Audiovisual Translation. The Encyclopedia of Applied Linguistics*. [http://www.researchschool.org/documents/Chiaro\\_Audiovisual%20Trl.pdf](http://www.researchschool.org/documents/Chiaro_Audiovisual%20Trl.pdf). (Dostop 31. 1. 2016)
- M. A. K. Halliday 1985. *An introduction to functional grammar*. London, Caulfield East, Baltimore. Edward Arnold.
- M. A. K. Halliday 1994. *An introduction to functional grammar, 2nd ed*. London. Arnold.
- Irena Kovačič. 1995. *Reinforcing or changing norms in subtitling*. V: C. Dollerup in V. Appel (eds) *Teaching Translation and Interpreting III*. Amsterdam. Benjamins, str. 105–109.
- Irena Kovačič. 1996. *Subtitling strategies: a flexible hierarchy of priorities*. V: C. Heiss in R.M. Bollettieri Bosinelli (eds) *Traduzione multimediale per il cinema, la televisione e la scena Bologna*, Clueb, str. 297–305.
- TED <http://www.ted.com/talks> (Dostop 17.1.2016)
- Darinka Verdonik. 2006. *Diskurzivni označevalci v telefonskih pogovorih*. Ljubljana. Slavistično društvo Slovenije.
- Urša Vogrinc Javoršek. 2007. *Metabesedilnost v okviru strategij podnaslovnega prevajanja*. Slavistično društvo Slovenije, letnik 52, številka 3/4, str. 79–94.