

Delavnica ELRC v Sloveniji

ELRC v Sloveniji

Simon Krek
Institut Jožef Stefan





- *ELRC-Share in slovenščina*
 - Simon Krek, , Institut “Jožef Stefan”
- *Projekt eTranslation TermBank*
 - Andraž Repar, Institut “Jožef Stefan”
- *Strojni prevajalnik Nematus @ IJS*
 - Matjaž Rihtar, Institut “Jožef Stefan”
- *Anonimizator Tacita*
 - Matej Kovačič, Institut “Jožef Stefan”



1. Bilingual text corpus

– Languages

- English (en)
- Slovenian (sl)

– Linguality

- Linguality type: Bilingual
- Multi-linguality type: Parallel

– Size

- **2,708,160** Words
- **55,184** Translation Units

– Text Format

- TMX

– Character encoding

- UTF-16LE

– Domains

- LAW (Eurovoc 12)



2. Bilingual text corpus

– Languages

- English (en)
- Slovenian (sl)

– Linguality

- Linguality type: Bilingual
- Multi-linguality type: Parallel

– Size

- **2,972,599** Words
- **63,069** Translation Units

– Text Format

- TMX

– Character encoding

- UTF-16LE

– Domains

- LAW (Eurovoc 12)



3. TERMIS: Slovene-English terminology in the field of public relations
 - Terminological Resource
 - Faculty of Social Sciences, Jožef Stefan Institute
4. Translations of Hungarian from public websites
 - A webcrawl of 14 different websites covering parallel corpora of Hungarian with Polish, Czech, Swedish, Finnish, French, German, Italian, English and Slovenian
 - Research Institute for Linguistics, Hungarian Academy of Sciences

5. University of Vienna Termbanks

- Attribution details: Department for Translation, University of Vienna
- 3 Termbanks about Risk Management, Austrian Asylum Law and University Law/Education Administration

6. Unofficial Consolidated legislative texts (Slovene)

- Attribution details: OPSI, Jožef Stefan Institute
- A collection (corpus in json format) of unofficial Consolidated text of the Laws, Regulations and other general acts in Slovenia.
 - 103,000,000 Tokens
 - 21,556 Files

National Anchor Points (NAP)

- NAP za javne storitve:
 - **Marčela Novljan Lovrinčič** (GSV, Služba za prevajanje)
- NAP za tehnologijo: **Simon Krek** (Institut Jožef Stefan)

Kdo je prispeval?

- GSV, Služba za prevajanje, OPSI

Kdo bi lahko prispeval?

- Kdorkoli s primernimi jezikovnimi viri!



Glavne ovire

Potrebe po tehnologiji, ki ni na voljo:
anonimizacija

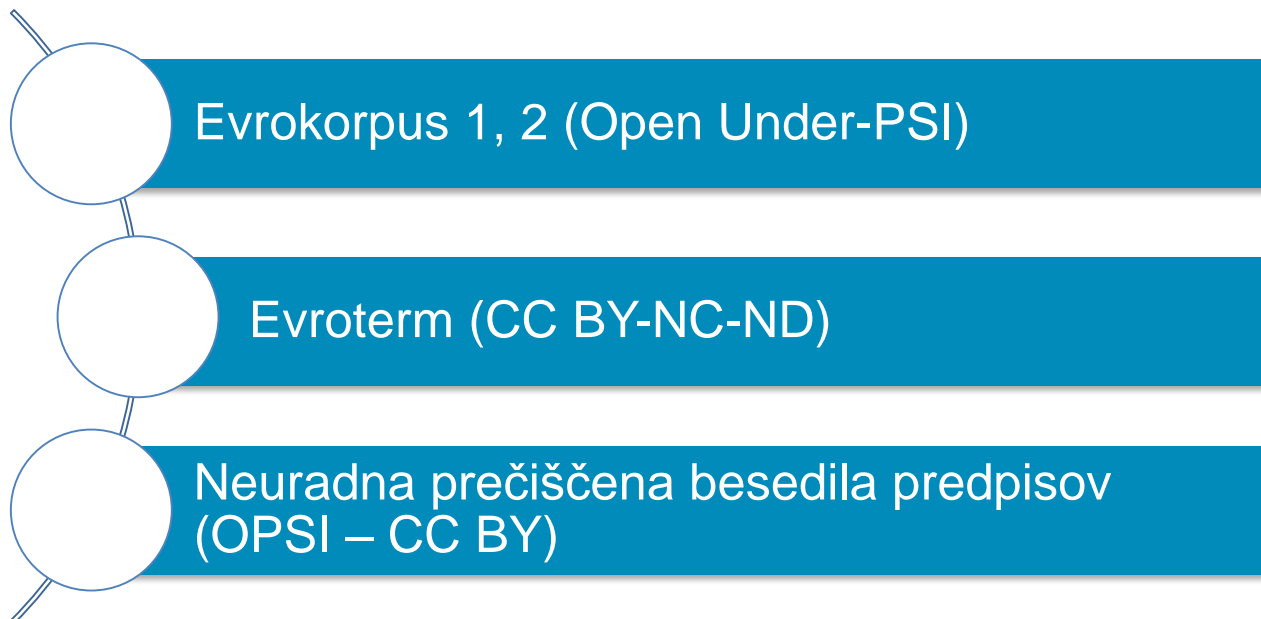
zadržki pravne narave (osebni podatki, lastništvo)

stranje gradiva:
pomnilniki prevodov?,
podvajanje besedil

Pomanjkanje časa za pripravo virov

Načelna zadržanost glede splošne dostopnosti virov

Glavni dosežki





- **8. december 2015**
- Panel 2: podatki in jezikovni viri v Sloveniji
- (testiranje MT@EC)
- Miran Željko
- (Generalni sekretariat vlade, Sektor za prevajanje)
- http://videolectures.net/elrc2015_zeljko_preizkus_prevajalnika/

- IZVIRNIK
 - **Public administrations in the European Union could benefit greatly by cooperating and exchanging know-how, experience and information.**
- MT@EC 2015
 - Javne uprave v Evropski uniji lahko velike koristi s sodelovanjem in izmenjavo znanja, izkušenj in informacij.
- IJS 2018
 - Javne uprave v Evropski uniji bi lahko v veliki meri imele koristi od sodelovanja in izmenjave informacij, izkušenj in informacij.

- **IZVIRNIK**
 - **If officials can access and share information across European borders, there is also potential for better and more efficient public administration.**
- **MT@EC 2015**
 - Če uradniki lahko dostop in izmenjavo informacij prek evropskih meja, obstajajo tudi možnosti za boljše in učinkovitejše javne uprave.
- **IJS 2018**
 - Če lahko uradniki dostopa in izmenjujejo informacije po evropskih mejah, obstaja tudi možnost za boljšo in učinkovitejše javno upravo.

- **IZVIRNIK**
 - **Imagine you're a busy employee in a public administration and receive a document in a foreign language.**
- **MT@EC 2015**
 - Veliko dela predstavljati zaposleni ste v javni upravi in prejem dokumenta v tujem jeziku.
- **IJS 2018**
 - Predstavljaš si, da si zaposlen v javni upravi in da imaš dokument v tujem jeziku.

- IZVIRNIK
 - **But you don't understand that language, can't wait for a human translator and aren't allowed to seek translation help over the regular Internet, for security reasons.**
- MT@EC 2015
 - Vendar ne razume, ne more čakati na prevajalca človeka in ne za prevajanje išče pomoč, redno internetu iz varnostnih razlogov.
- IJS 2018
 - Ne razumeš tega jezika, ne moreš čakati na prevajalca in ne smeš iskati prevodov, ki bi nam pomagali preko rednega interneta, zaradi varnostnih razlogov.

- IZVIRNIK
 - **Needless to say, Mr Hilary is entitled to his view.**
- MT@EC 2015
 - Samo po sebi je umevno, da je upravičen do Hilaryju svoja stališča.
- IJS 2018
 - Odveč je reči, da ima gospod Hilary pravico do svojega mnenja.



- *ELRC-Share in slovenščina*
 - Simon Krek, , Institut “Jožef Stefan”
- *Projekt eTranslation TermBank*
 - Andraž Repar, Institut “Jožef Stefan”
- *Strojni prevajalnik Nematus @ IJS*
 - Matjaž Rihtar, Institut “Jožef Stefan”
- *Anonimizator Tacita*
 - Matej Kovačič, Institut “Jožef Stefan”