

Kakšne podatke potrebujemo

RED. PROF. DR. ŠPELA VINTAR











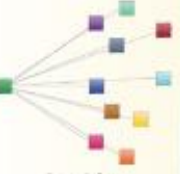




ODDELEK ZA PREVAJALSTVO

FILOZOFSKA FAKULTETA

UNIVERZA V LJUBLJANI

Translation Industry Evolution

Translation to Convergence

	1980 TRANSLATION	1990 LOCALIZATION	2000 GLOBALIZATION	2010 INTEGRATION	2020 CONVERGENCE
TRANSLATION INDUSTRY FOCUS	 Documents	 Software	 Simship	 Integration in enterprise systems	 Embedded in every app, on every screen
TYPE OF CONTENT TRANSLATED	 Paper	 Digital	 Static Web	 Dynamic Web	 Personalized
LANGUAGES	 1 > 10	 1 > 25	 1 > 40	 6 <> 60	 150 <> 150
TRANSLATION SUPPORT					
TOOLS	NA	TM and terminology software	Workflow (GMS)	MT and Advanced Leveraging	Real-time customized MT
RESOURCES	Glossaries	Project translation memories	Centralized corporate TMs	Limited shared data	Unlimited big data

Source: TAUS

Profesionalno prevajanje pri nas

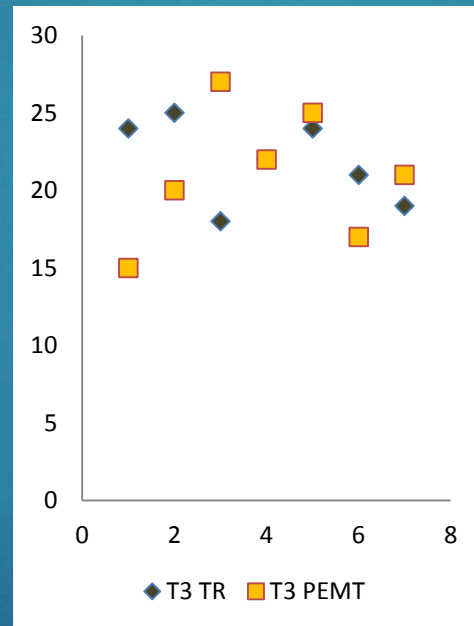
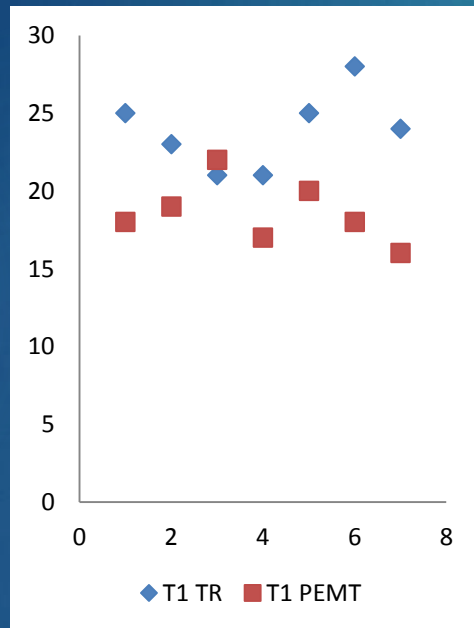
- ▶ **Povpraševanje** po prevajalskih storitvah v svetu in pri nas narašča.
- ▶ 93,4 % profesionalnih prevajalcev pozna orodja za **računalniško podprto prevajanje**, 74,4 % jih redno uporablja (Žunič 2013).
- ▶ Večje prevajalske agencije v delovni proces uvajajo **strojno prevajanje** in s tem povečajo produktivnost (Artiček 2014).
- ▶ **Popravljanje strojnih prevodov** predstavlja vse pomembnejši delež prevajalskega dela.
- ▶ Popravljanje strojnih prevodov je **hitrejše** od prevajanja od začetka, tudi če je uporabljen splošni prevajalnik in če popraviljalci za to nalogo niso posebej usposobljeni (Smolič 2015).

Študija primera: agencija si kupi prevajalnik

- ▶ izbrani ponudnik: Asia Online, jezikovna para AN-SL in SL-AN
- ▶ začetna investicija: 40.000 \$, nato obračunavanje po obsegu (0,008 \$/besedo)
- ▶ ponudnik sam zagotovi splošne vire, agencija prispeva lastne pomnilnike, glosarje, skenirane strokovne slovarje in kakovostne enojezične vire
- ▶ začetno urjenje prevajalnika traja 3 mesece
- ▶ končni prevajalnik v avtomatski evalvaciji (BLEU) preseže Google in Bing:
 - ▶ agencijski prevajalnik 25,62, Google 21,60 in Bing 15,12.

(vir: Artiček 2014)

Študija primera II: Študenti popravljajo strojni prevod



- ▶ povprečni čas PEMT je krajši od T (razlika je majhna)
- ▶ anketa: 66 % študentov ocenjuje, da bi se lahko navadili na PEMT
- ▶ 85 % študentov meni, da je prevod od začetka bolj kakovosten od popravljenega strojnega prevoda

CAT in MT



samodejni
slovarji fraz



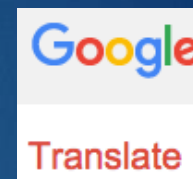
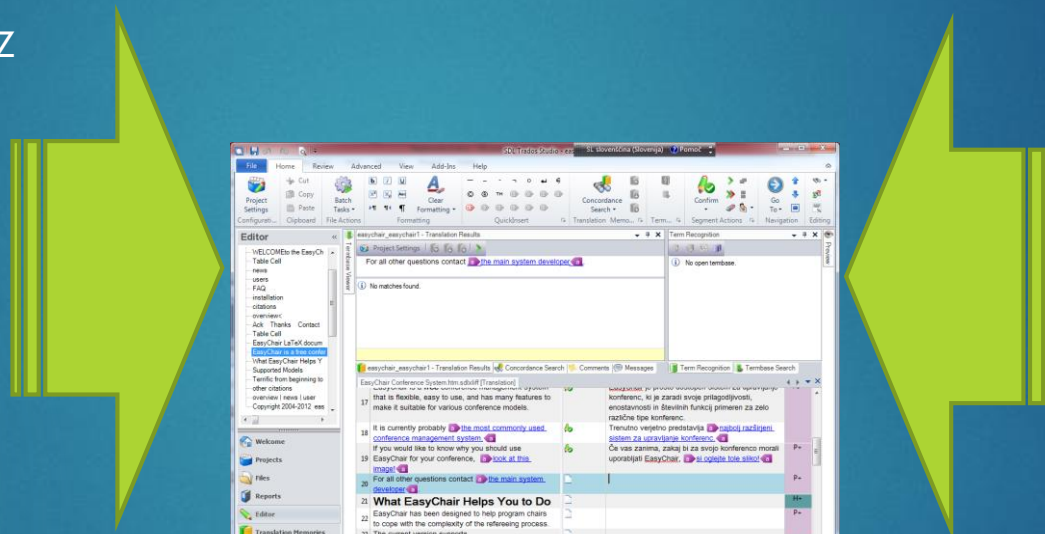
pomnilnik
prevodov



term.
baza



arhivi
prevodov



EVROTERM



Installed machine translation plugins



Crosslang Gateway MT

Copyright 2013 CrossLang NV

Options

Enable plugin



GoogleMT machine translation plugin

(C) Kilgray 2009

Options

Enable plugin



IPTranslator MT plugin

(C) Iconic Translation Machines Ltd.

Options

Enable plugin



Microsoft MT plugin

(C) Kilgray

Options

Enable plugin



iTranslate4.eu MT plugin

(C) Kilgray

Options

Enable plugin



Pangea MT plugin

(C) Pangeanic

Options

Enable plugin



KantanMT API Connector

(C) Xcelerator Machine Translations Ltd. 2014

Options

Enable plugin

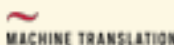


memoQ pseudo-translation plugin

(C) Kilgray

Options

Enable plugin



Let'sMT! plugin

(C) Moravia Worldwide 2012

Options

Enable plugin



Systran MT plugin

(C) Kilgray

Options











Enable plugin

Let'sMT

Systems

▼ Status ▼ Source language ▼ Target Language ▼ Subject domain

-select- English Slovenian -select-

Title	Language	Domain	Created
 HIPRE	🌐 English - Slovenian	📄 Law	2015.08.29  Public
 [TransTech 2015 English to Slovenian] Audiovisual Translation (subtitles)	🌐 English - Slovenian		2015.06.18  Public
 English - Slovenian IT IT domain engine	🌐 English - Slovenian	📄 IT	2014.10.10  Public
 English - Slovenian Legal Legal domain engine	🌐 English - Slovenian	📄 Legal	2014.10.07  Public
 English - Slovenian Pharmaceutical English - Slovenian Pharmaceutical engine	🌐 English - Slovenian	📄 Pharmaceutical	2014.09.28  Public

Množičenje prevodov

▶ plačani prevodi množic

- ▶ do 80 % cenejši
- ▶ popravljanje strojnih prevodov
- ▶ za boljšo kakovost dvojno preverjanje + preverjanje kompetenc z drugimi nalogami

▶ znane platforme:

- ▶ Ackuna, <http://ackuna.com/>
- ▶ Gengo, <http://gengo.com/>
- ▶ Translation Cloud, <http://www.translationcloud.net/>

▶ prostovoljski prevodi

- ▶ odprtnokodna skupnost
- ▶ “entuziasti” -> občasno celo boljši!
- ▶ številne znane aplikacije
 - ▶ Mozilla, Linux, Wordpress, OpenOffice, LibreOffice ...
 - ▶ TED Open Translation Project, <https://www.ted.com/participate/translate/guidelines>



You like this.

Sentence on an object listing the people who like it.

Všeč ti je.

To tje full kull! :-D

To vam je všeč.

zaradi tega švicaš fuž.

To ti je všeč.

Tole tje kul ...

+ New translation

- ▶ Storitve strojnega prevajanja so **plačljive**...
- ▶ ...čeprav v veliki meri temeljijo na **javno dostopnih podatkih**.
- ▶ Za prevajalce in prevajalnike so najdragocenejši **pomnilniki prevodov**, ti pa so večinoma v lasti korporacij.
- ▶ Tudi prevodni podatki javnih institucij **niso javno dostopni**.
- ▶ Korporativne terminološke baze skoraj nikoli **niso javno dostopne**.
 - ▶ ...s svetlo izjemo Microsofta.
- ▶ Tudi terminološke baze javnih institucij **niso javno dostopne**.
 - ▶ ...vsaj ne v uporabnem formatu (TBX).
- ▶ Pravni vidiki pravic za prevodne podatke **niso povsem jasni**.

Kdo je lastnik pomnilnika prevodov?

- ▶ Prevod
 - ▶ je avtorsko delo in je avtorskoppravno zaščiten. (Izjema so uradne objave, sodna in upravna besedila.)
 - ▶ Moralnih avtorskih pravic prevajalca ni mogoče prenesti na drugo osebo.
 - ▶ Materialne avtorske pravice praviloma pripadajo naročniku ali – v primeru delovnega razmerja – delodajalcu.
- ▶ Pomnilnik prevodov
 - ▶ je podatkovna zbirka kratkih segmentov besedila, ki tipično vsebuje podatke iz več virov.
 - ▶ ni avtorsko delo.
 - ▶ je intelektualna lastnina ponudnika prevajalskih storitev.
 - ▶ Če pomnilnik prevodov omogoča rekonstrukcijo (in redistribucijo) celotnega besedila, je to kršitev pravic intelektualne lastnine.
 - ▶ (...še vedno pa se pravice do pomnilniških in drugih izpeljanih baz določajo od primera do primera).
- ▶ Terminološka baza
 - ▶ načeloma pripada tistemu, ki jo je ustvaril (in/ali naročil).
 - ▶ Jezikovni podatki v njej ne morejo biti predmet avtorskoppravne zaščite, razen če terminološka baza vsebuje daljše citate ali odseke besedila.

- ▶ Clarifying Copyright on Translation Data, TAUS, <https://www.taus.net/think-tank/articles/translate-articles/clarifying-copyright-on-translation-data>
- ▶ Smith, R. (2009) Copyright Issues in Translation Memory Ownership. ASLIB. (pdf v e-učilnici)
- ▶ Translation Memory (TM) is the translator's Intellectual Property (IP), ProZ.com, http://www.proz.com/forum/business_issues/287990-translation-memory-tm-is-the-translators-intellectual-property-ip.html
- ▶ Rob Munro: Translation and Crowdsourcing, <http://idibon.com/translation-and-crowdsourcing/>
- ▶ What is the most successful translation crowdsourcing project and why? <https://www.quora.com/What-is-the-most-successful-translation-crowdsourcing-project-and-why>
- ▶ Artiček, Tadeja (2014) Uvajanje strojnega prevajanja v delo prevajalske agencije : analiza popravljanja strojnih prevodov. Magistrsko delo. Univerza v Ljubljani, Filozofska fakulteta. Mentorica Š. Vintar.
- ▶ Smolič, Sergej (2015) Značilnosti popravljanja strojnih prevodov v primerjavi s klasičnim prevajanjem. Magistrsko delo. Univerza v Ljubljani, Filozofska fakulteta. Mentorica Š. Vintar.
- ▶ Žunič, Katja (2013) Pomen računalniško podprtega prevajanja za študente prevajalstva in prevajalce v prevajalskih agencijah. Diplomsko delo. Univerza v Mariboru, Filozofska fakulteta. Mentorica V. Kučič.
- ▶ TAUS Translation Technology Landscape Report, 2013, <https://www.taus.net/think-tank/reports/translate-reports/taus-translation-technology-landscape-report>