

# **LSP & POST-EDITING OF MACHINE TRANSLATION**



# Basic terms

- LSP - **Language for Specific Purposes** (technical language)
- LSP is a language used in a specific subject field -> field of specialized knowledge
- A **term** is a designation of a concept belonging to a language for specific purposes (*termin/naziv – terminologija/nazivlje*)



# Basic terms

- A term designates a **concept** (*pojam*) -> hurricane, cyclonic storm, flood, levee, fire hose, crisis management, myopia, presbyopia, keratoconus, cornea...
- We **define** terms according to their position and relation within a **specialized domain** -> crisis management, optometry, law, business, medicine...
- **Terminology** - discipline which studies terms and their usage and application within a domain.



# Usage

- Reading and understanding **technical texts**.
- Using appropriate terms when writing technical texts.
- **Translation of technical content** (work e-mails, thesis abstracts).



# Lexical resources

- **DICTIONARY** - a book or electronic resource that lists the words of a language in alphabetical order and gives their meaning
- **GLOSSARY** - an alphabetical list of words relating to a specific subject, text, or dialect, with explanations; a short dictionary
- **TERMBASE** - a database consisting of concept-oriented terminological entries (concepts) and related information



**MONOLINGUAL**



**BILINGUAL**



**MULTILINGUAL**



word

pronunciation

part of speech

first definition

sample sentence

**volcano** (väl kā'nō') *noun* **1.** a vent in Earth's crust through which molten lava and gases are ejected; Red Mountain, in northern Arizona, is one of many mountains that formed as the result of a *volcano*. **2.** something with explosive potential; Before World War II, Germany was a *volcano* of unrest.

sample sentence

second definition



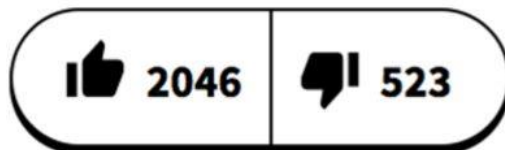
# Selfie

The beginning of the end of intelligent civilization

*Future sociologists use the selfie as an artifact for the end of times*

#picture #photograph #groupies #selfy #celluloid

by **Future Sociologist** November 27, 2013







Projekti

O Struni

Publikacije

Novosti

ACTT 2013.

Poveznice

Kontakt



Pomoć ?

+ Napredno pretraživanje



terminološka baza podataka

TRAŽI

© 2011 Institut za hrvatski jezik i jezikoslovlje

© Srećko Niketić / Cropix

Rezultati pretraživanja za: **TERMINOLOŠKA BAZA PODATAKA**

terminološka baza podataka

organizirana zbirka pretraživih terminoloških podataka pohranjenih u elektroničkome obliku

filologija

# Bilingual glossaries

**SOURCE  
TERM**

**TARGET  
TERM**

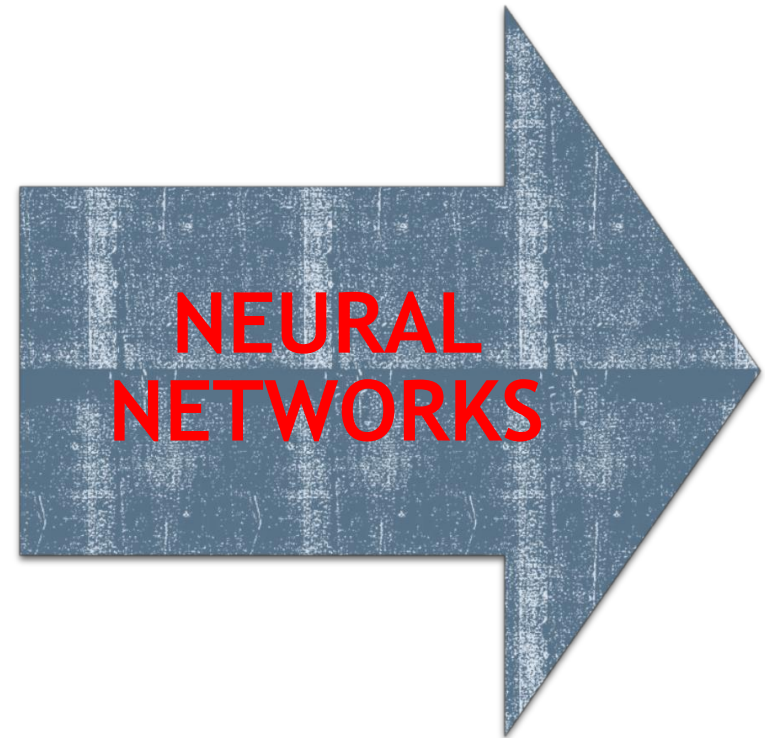
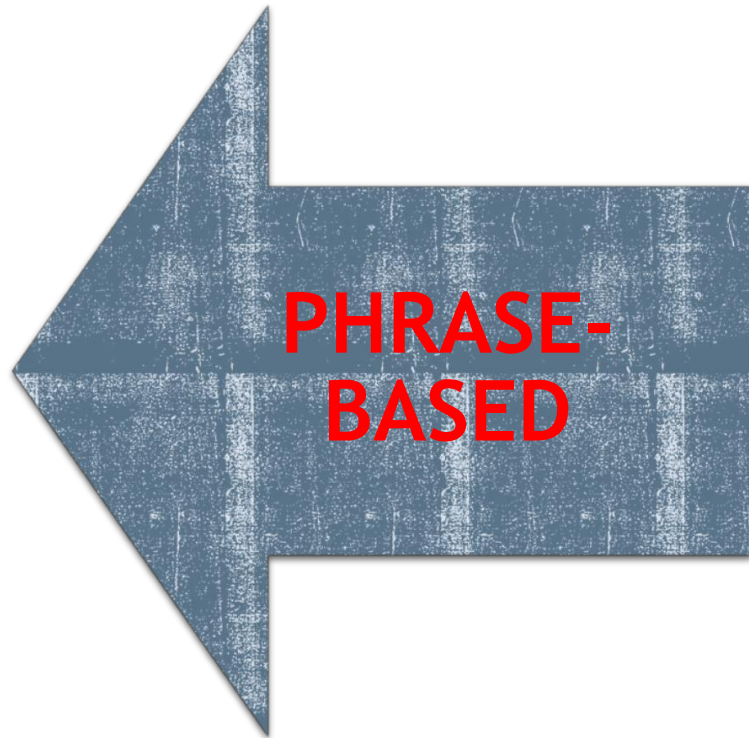


# Machine translation

- There are several types of MT:
- Statistical Machine Translation (SMT)
- Rule-based Machine Translation (RBMT)
- Phrase-based Machine Translation (PBMT)
- Neural Machine Translation



# Human translation vs. machine translation



# Machine translation

- **Neural machine translation** teaches itself on how to translate by using a large **neural network**.
- Pitfalls: **mistranslations while the target text looks very natural** - hence it's harder to spot issues when post-editing.
- **Document-level context**, as well as the **socio-cultural nuances**, remain a challenge to all MT engines, for the time being (e. g. not differentiating between *he* or *she*).
- **Low-resource languages**: pivot translation.



MEDIA AND TELECOMS JANUARY 18, 2020 / 6:27 PM / UPDATED 3 YEARS AGO

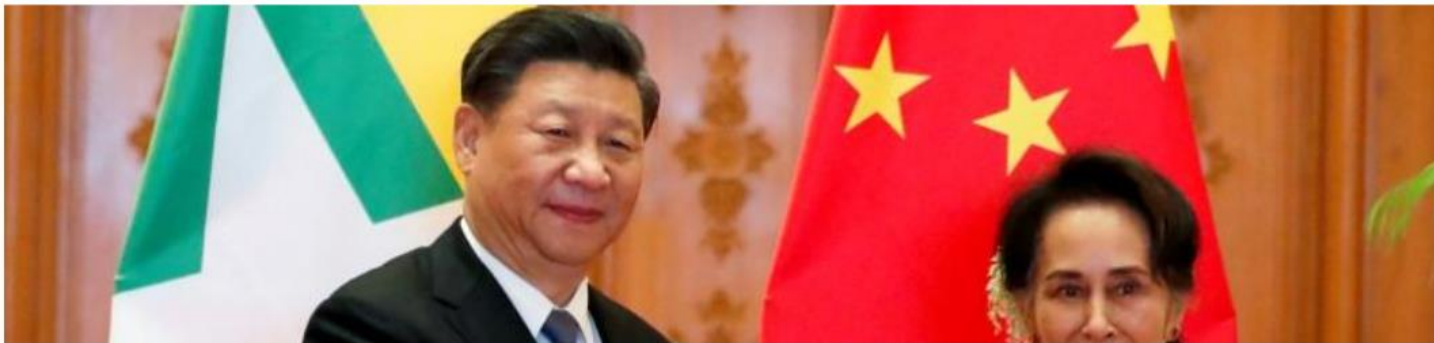
## Facebook says technical error caused vulgar translation of Chinese leader's name

By Poppy McPherson

3 MIN READ



YANGON (Reuters) - Facebook Inc [FB.O](#) on Saturday blamed a technical error for Chinese leader Xi Jinping's name appearing as "Mr Shithole" in posts on its platform when translated into English from Burmese, apologizing for any offense caused.





# Machine translation

- Large **corpora** on which neural networks are **trained** (corpora of texts, databases, the Internet).
- Various MT programs: **Google Translate**, **DeepL**, **Systran**, **Baidu Translate**, **Tencent TranSmart**, **eTranslation** used by the European Commission.
- **Deep learning** -> **machine learning** methods based on **learning data representations**, as opposed to task-specific algorithms.



# Machine translation & ChatGPT

- ChatGPT is a chatbot developed by OpenAI.
- It has been trained to follow and answer instructions or prompts written by users.
- Jiao et. al. (2022): ChatGPT produces similar results to standard online MT systems for texts that are like its training data (user-generated social media posts in high-resource languages such as English and German).
- It produces worse results for low-resource languages such as Romanian or languages distant from English, such as Japanese.





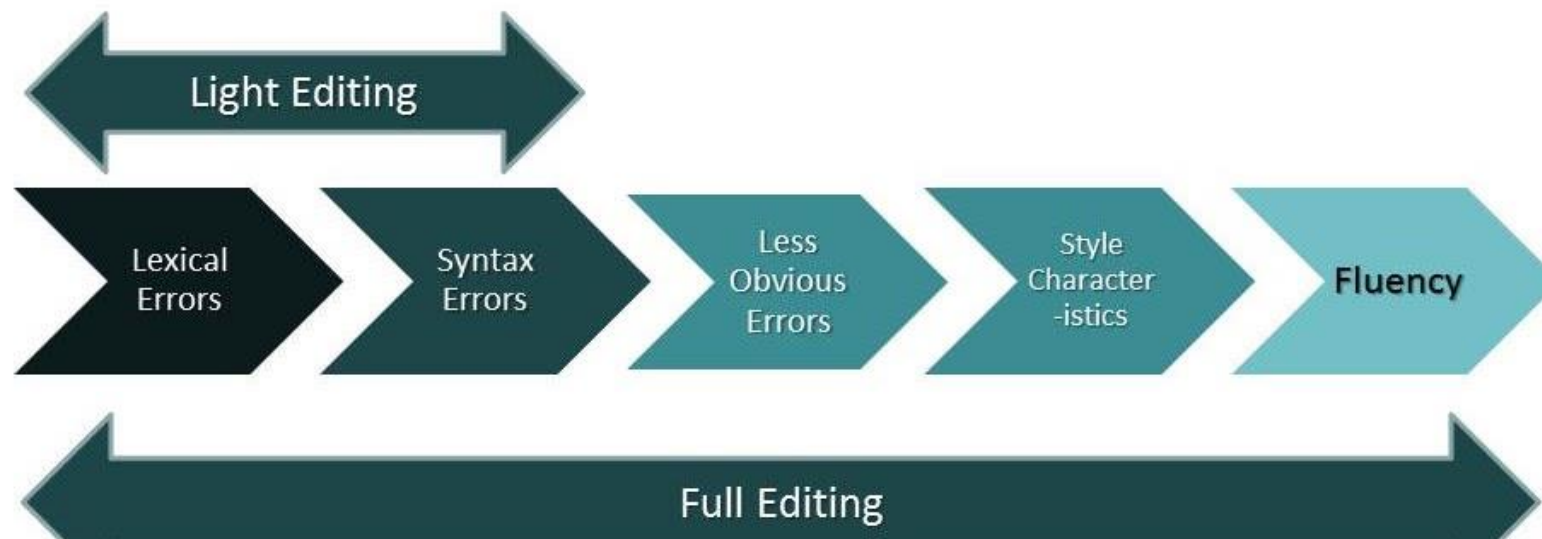
# QUALITY STANDARDS

- Human translation quality standard: ISO 17100:2015
- Machine translation quality standard: ISO 18587:2017
- ISO 18587:2017 provides requirements for the process of full, human post-editing of machine translation output and post-editors' competences.
- Essentially: a post-editor is expected to do what he or she would normally do while translating or revising a human translation.



- Human translation: translation + revision (self-revision, other-revision)
- Machine translation: translation + post-editing (revision of a machine translated text)

## WHAT IS POST-EDITING? THE LARGER CONTEXT



# Competences of a post-editor

- Good knowledge of the:
  - **source language**
  - **target language**
  - **domain-specific terminology**
- + **Ability to explain and bolster changes made to the text.**



# Pre-editing for MT

- The source text should be carefully formatted and **pre-edited**.
- One **typo** can mess up the entire paragraph.
- The main steps to get your content ready for translation include ensuring **consistent terminology use** and formatting sentences to be **straightforward** and **uncomplicated**.
- Source text should be as **clear and concise** as possible. The simpler the original input, the better the end results.



# Pre-editing for MT

- Machine engines tend to have problems with **sarcasm**, **double-negatives**, and **words with several meanings**.
- **Idiomatic language** is still a problem for MT.
- For best results: **simple grammar**, **short sentences**, **unambiguous phrases**, **subject-verb-object** sentence structure.
- NMT is still **as good as its training databases** (low-resource languages such as Croatian could generate more error-prone results).



# Typical errors

- **Mistranslations** of words, phrases, proper names etc.
- **Negation and opposite meaning errors** refer to mistakes in the target language that result in the opposite meaning of what was intended by the original text.
- Sometimes MT can introduce content that is simply not present in the source. These are called **hallucinations** (usually a problem with MT engine itself).



# Post-editing guidelines

Lexical and syntactic level	Terminology usage	Stylistic level	References
Grammar, spelling and punctuation must be correct.	Correct terminology must be used & used consistently.	Use of appropriate style & register and a translation that is suitable for its purpose.	The same references are used as for human revision: dictionaries, lexicons, glossaries, termbases, style guides.



# **POST-EDITING OF A LSP TEXT – PRACTICAL EXAMPLES**







## Translate

Turn off instant translation

Croatian English Spanish Detect language



English Croatian Spanish

Translate

Mrežnica je funkcionalno najvažniji dio oka, te se može usporediti s filmom u fotoaparatu. Uloga mrežnice je primati svjetlosne podražaje, koje mijenja kompliciranim kemijskim procesom u električne impulse vođene dalje prema mozgu. Veličina zjenice određuje količinu svjetla koje ulazi u oko. Zjenični otvor povezuje prednju i stražnju očnu sobicu, te kroz njega cirkulira očna vodica.



384/5000



The mesh is functionally the most important part of the eye, and can be compared to the film in the camera. The role of the retina is to receive light stimuli, which are transformed by complicated chemical processes into electrical impulses driven further to the brain. The pupil's volume determines the amount of light entering the eye. The eye opening connects the front and back ocular chambers, and the occiput circulates through it.



Suggest an edit

Google Translate

https://translate.google.com/#hr/en/Mrežnica je funkcionalno najvažniji dio oka%20

Translate

Turn off instant translation

Croatian English Spanish Detect language

English Croatian Spanish Translate

Mrežnica je funkcionalno najvažniji dio oka, te se može usporediti s filmom u fotoaparatu. Uloga mrežnice je primati svjetlosne podražaje, koje mijenja kompliciranim kemijskim procesom u električne impulse vođene dalje prema mozgu. Veličina zjenice određuje količinu svjetla koje ulazi u oko. Zjenični otvor povezuje prednju i stražnju očnu sobicu, te kroz njega cirkulira očna vodica.

384/5000

The mesh is functionally the most important part of the eye, and can be compared to the film in the camera. The role of the retina is to receive light stimuli, which are transformed by complicated chemical processes into electrical impulses driven further to the brain. The pupil's volume determines the amount of light entering the eye. The eye opening connects the front and back ocular chambers, and the occiput circulates through it.

Suggest an edit

Mrežnica - retina  
the camera - a camera  
driven - led  
further - then  
volume - size  
pupillary opening/aperture

anterior and posterior ocular chamber  
očna vodica - aqueous humour

About Google Translate Community Mobile G+ B

About Google Privacy & Terms Help Send feedback

14:29 8.10.2018.



Sign in

Translate

Turn off instant translation



Croatian English Spanish Detect language ▾



English Croatian Spanish ▾

Translate

In western countries, rivers prone to flooding are often carefully managed. Defences such as levees, bunds, reservoirs, and weirs are used to prevent rivers from bursting their banks. Coastal flooding has been addressed in Europe with coastal defenses, such as sea walls and beach nourishment. ✕



293/5000

U zapadnim zemljama rijeke sklone poplavama često se pažljivo upravljaju. Obrane kao što su kapljevine, bundovi, rezervoari i čvorovi koriste se kako bi se spriječilo da rijeke rastu iz svojih banaka. Obalna poplava obrađena je u Europi s obalnim obranama, kao što su morske zidine i plaže.



Suggest an edit

About Google Translate

Community

Mobile



About Google

Privacy & Terms

Help

Send feedback



HRV

23:53  
14.10.2018.



Google Translate

https://translate.google.com/#en/hr/In western countries%2C rivers prone to

Translate

Turn off instant translation

Croatian English Spanish Detect language

English Croatian Spanish Translate

In western countries, rivers prone to flooding are often carefully managed. Defences such as levees, bunds, reservoirs, and weirs are used to prevent rivers from bursting their banks. Coastal flooding has been addressed in Europe with coastal defenses, such as sea walls and beach nourishment.

U zapadnim zemljama rijeke sklone poplavama često se pažljivo upravljaju. Obrane kao što su kapljevine, bundovi, rezervoari i čvorovi koriste se kako bi se spriječilo da rijeke rastu iz svojih banaka. Obalna poplava obrađena je u Europi s obalnim obranama, kao što su morske zidine i plaže.

Rijekama se pažljivo upravlja.

levee - nasip  
bund - zečji nasip  
weir - ustava/brana  
(river)bank - riječna obala

to address something -  
**pozabaviti se/pobrinuti se**  
coastal defenses - **obalni obrambeni sustavi**  
sea wall - **valobran/lukobran**

About Google Translate Community Mobile G+ B

About Google Privacy & Terms Help Send feedback

23:53 14.10.2018.

Text Images Documents Websites

DETECT LANGUAGE

CROATIAN

ENGLISH

GERMAN



ENGLISH

CROATIAN

SPANISH

BSE is thought to have come from a similar disease in sheep called scrapie. In the 1980s, producers of cattle feed (which often included ground meat and bone meal by-products from sheep) changed the way they processed feed. The change somehow allowed the scrapie disease agent to survive the cattle feed production process. Thus, contaminated food was fed to cattle, which then came down with BSE. At the time, neither scrapie nor BSE were thought to affect humans. So, meat (nervous tissue) from BSE-infected cows made it into the food supply. Humans who ate the infected meat (probably hamburger or other processed meats) contracted the BSE-causing agent and developed nvCJD.



677 / 5,000



Smatra se da BSE dolazi od slične bolesti kod ovaca koja se zove scrapie. U 1980-ima proizvođači stočne hrane (koja je često uključivala mljeveno meso i nusproizvode od koštanog brašna ovaca) promijenili su način na koji su prerađivali stočnu hranu. Promjena je nekako omogućila uzročniku bolesti scrapie da preživi proces proizvodnje stočne hrane. Tako je kontaminiranom hranom hranjena stoka, koja je zatim oboljela od GSE-a. U to se vrijeme smatralo da ni scrapie ni BSE ne mogu utjecati na ljude. Tako je meso (živčano tkivo) krava zaraženih BSE-om dospjelo u zalihe hrane. Ljudi koji su jeli zaraženo meso (vjerojatno hamburger ili drugo prerađeno meso) dobili su uzročnika BSE-a i razvili nvCJD.



Send feedback

Text Images Documents Websites

DETECT LANGUAGE

CROATIAN

ENGLISH

GERMAN



ENGLISH

CROATIAN

SPANISH

BSE is thought to have come from a similar disease in sheep called scrapie. In the 1980s, producers of cattle feed (which often included ground meat and bone meal by-products from sheep) changed the way they processed feed. The change somehow allowed the scrapie disease agent to survive the cattle feed production process. Thus, contaminated food was fed to cattle, which then came down with BSE. At the time, neither scrapie nor BSE were thought to affect humans. So, meat (nervous tissue) from BSE-infected cows made it into the food supply. Humans who ate the infected meat (probably hamburger or other processed meats) contracted the BSE-causing agent and developed nvCJD.



677 / 5,000



Smatra se da BSE dolazi od slične bolesti kod ovaca koja se zove scrapie. U 1980-ima proizvođači stočne hrane (koja je često uključivala mljeveno meso i nusproizvode od koštanog brašna ovaca) promijenili su način na koji su prerađivali stočnu hranu. Promjena je nekako omogućila uzročniku bolesti scrapie da preživi proces proizvodnje stočne hrane. Tako je kontaminiranom hranom hranjena stoka, koja je zatim oboljela od GSE-a. U to se vrijeme smatralo da ni scrapie ni BSE ne mogu utjecati na ljude. Tako je meso (živčano tkivo) krava zaraženih BSE-om dospjelo u zalihe hrane. Ljudi koji su jeli zaraženo meso (vjerojatno hamburger ili drugo prerađeno meso) dobili su uzročnika BSE-a i razvili nvCJD.

Send feedback

**BSE - BOVINE SPONGIFORM ENCEPHALOPATHY (KRAVLJA SPONGIFORMNA ENCEFALOPATIJA - KRAVLJE LUDILO)**

**SCRAPIE - GREBEŽ OVACA**

**nvCJD - NEW VARIANT CREUTZFELDT-JAKOB DISEASE (NOVA VARIJANTA CREUTZFELDT-JAKOBOVE BOLESTI - LJUDSKI OBLIK KRAVLJEG LUDILA)**

# Bottom line

- **NEVER** copy-paste an entire text into a MT tool.
- **PREPARE** your text before you use MT.
- **ALWAYS** read the MT output after you have used MT.
- If necessary, **POST-EDIT** the output (bearing in mind the guidelines for post-editing).



**THANK YOU FOR ATTENTION!**

