Machine Translation

Radhika Mamidi

Little bit of warm-up…

Exercise 1: Translate the following sentences in your mother tongue and comment.

- I went to the shop and bought a few things that I love.
- Please print the poem which was written by Tagore.
- My aunt, who lives in Malaysia, brought this book on Gandhi.
- My cousins love cooking. They download recipes and make new dishes every evening. They are delicious.
- She is my sister. She loves reading books. This is her book.

Exercise 2: Translate the following sentences in your mother tongue and comment.

- John killed Bill. Bill died.
- Harry made John kill Bill.
- George swam to the other end.
- Mary drives Anne to school every day.
- Jenny broke the window open.

TRANSLATION IS THE NEED OF THE DAY

Why translation?

- We live in an era of rapid globalization, which is demonstrated by the growing demand for language services.
- Translation is important because it facilitates multilingual communication and allows people from around the world to better understand one another culturally, economically and socially.
- Translation technology increase human translator productivity by up to 400%.

Human Translators

- High quality
- Publishable translation

"Despite all the recent successes of MT, when it comes to high quality publishable translation, human translators are still unchallenged".

Reference: Koehn, Philipp. Computer Aided Translation: MSR talk, 9 November 2012.

Limitations of Human Translators

- Human translators can be slow and sloppy
- We waste time searching for common vocabulary we have forgotten, and a lot of time is spent looking things up.

Lost in translation

- In a Tokyo bar:
  Special cocktails for the ladies with nuts.
- On the menu of a Swiss restaurant:
  Our wines leave you nothing to hope for.
- Hotel lobby, Bucharest:
  The lift is being fixed for the next day. During that time we regret that you will be unbearable.
- Outside a dress shop, Hong Kong:
  Ladies have fits upstairs.
- Hotel elevator, Paris:
  Please leave your values at the front desk.

Machine Translation

Machine Translation (MT) is a subfield in computational linguistics where computer programs are used for automatically translating one natural language to another.

Advantages of MT

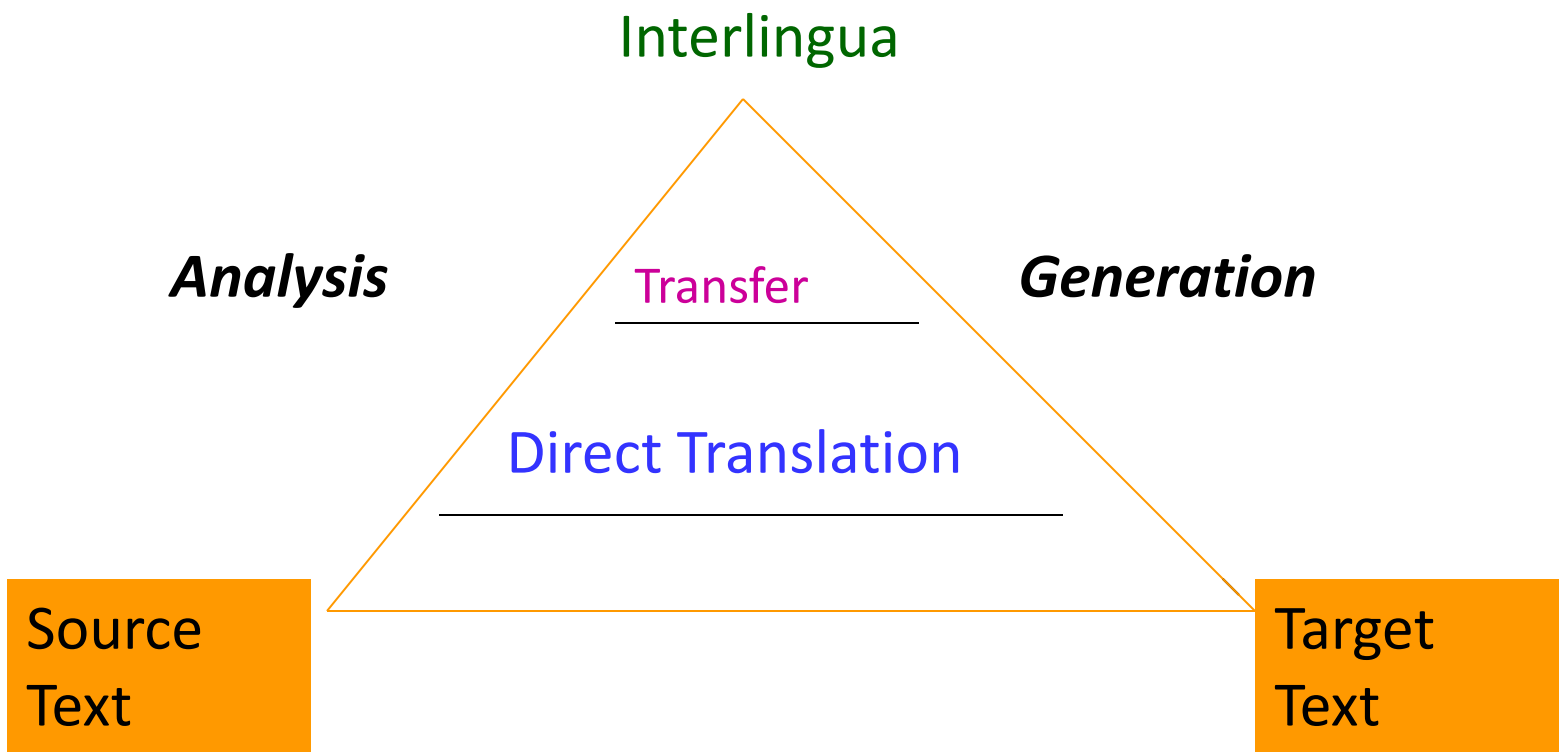
- Fast
- Comparatively cheap
- Confidential

Very useful for…

- Translating user manuals
- Translating UN, EU documents
- Translating instructions
- Translating web pages

- Successful in domain based translation – medical, sports, science, technology etc.

Approaches


Interlingua
Analysis
Transfer
Generation
Direct Translation
Source Text
Target Text

Methods

1. Rule based:

- Words will be translated in a linguistic way — the most suitable words of the target language will replace the ones in the source language
- This method requires extensive lexicons with morphological, syntactic, and semantic information, and large sets of rules.

Methods

2. Corpus based:

- Usually statistics is used to find the equivalents in two languages from the parallel corpus for a given word, phrase or sentence.
- Given enough data, machine translation programs often work well enough for a native speaker of one language to get the approximate meaning of what is written by the other native speaker.
- Alignment in parallel corpus

Types of MT systems

1. **Dictionary based MT**

   Machine translation can use a method based on dictionary entries, which means that the words will be translated as a dictionary does — word by word.

Types of MT systems

2. Statistical machine translation

The machine translation tries to generate translations using statistical methods based on bilingual parallel corpora. The machine learns from the aligned parallel corpora.

3. Example-based machine translation

The EBMT approach is often characterised by its use of a bilingual corpus as its main knowledge base, at run-time. It is essentially a translation by analogy.

NMT

4. **Neural machine translation (NMT)**

It is an approach to machine translation that uses an artificial neural network. It predicts the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.

They require only a fraction of the memory needed by traditional statistical machine translation (SMT) models. Furthermore, unlike conventional translation systems, all parts of the neural translation model are trained jointly (end-to-end) to maximize the translation performance.

(Wikipedia)

Best example is of Google Translate service which adopted NN approach in 2016. It was an SMT service earlier.

Evaluation

- Evaluating the MT system is essential.
- There are various methods for evaluating the performance of machine translation systems:
  - The oldest method is by using human judges to tell the quality of a translation,
  - Automated methods include BLEU, NIST and METEOR.

Some issues

- Languages have different word orders/structures.
- Some languages are morphologically complex.
- Some languages do not have determiners.
- Identifying and finding equivalents of idioms, collocations, phrasal verbs etc.
- Lexical gaps
- Lexical ambiguity
- Structural ambiguity

The challenge

- How to program a computer to "understand" a text as a human being does
- Also to "create" a new text in the source language that "sounds" as if it has been written by a human

Limitations of MT

Translation software gives the best approximation of the source text but it has a lot of limitations:

➔It can’t translate cultural components of the source text into the target language;

➔The translation is very literal, often word to word;

➔It doesn’t recognize idioms, slang, and terms that are not in the machine’s memory;

➔It lacks the creativity of the human touch.

MT: Aiding Human Translators

- MT output is a good draft for post-editing by human translators.
- MT can't beat human translators but it can boost their productivity.

Compromise

- It would be absurd to claim that a machine could produce a target text of the same quality as that of a human being
- A machine can do the first stage of translation automatically and then human beings can revise and edit the translation.
  - Machine/Computer Aided Translation systems
  - Human Aided Machine Translation systems

File Edit View History Bookmarks Tools Help

Zimbra: Search results | Google Translate

https://translate.google.co.in/#auto/te/One day I will go to Disneyland. One day we will go to Disneyland. One day she

story tortoise and hare

Firefox has prevented the outdated plugin "Adobe Flash" from running on translate.google.co.in.

Try a new browser with automatic translation. Download Google Chrome Dismiss

Google

Translate

English Spanish French English - detected

English Spanish Telugu Translate

One day I will go to Disneyland. One day we will go to Disneyland. One day she will go to Disneyland.

ఒక రోజు నేను డిస్నీల్యాండ్ వెళతారు. ఒక రోజు వె
వెళతారు.

Type text or a website address or translate a document.

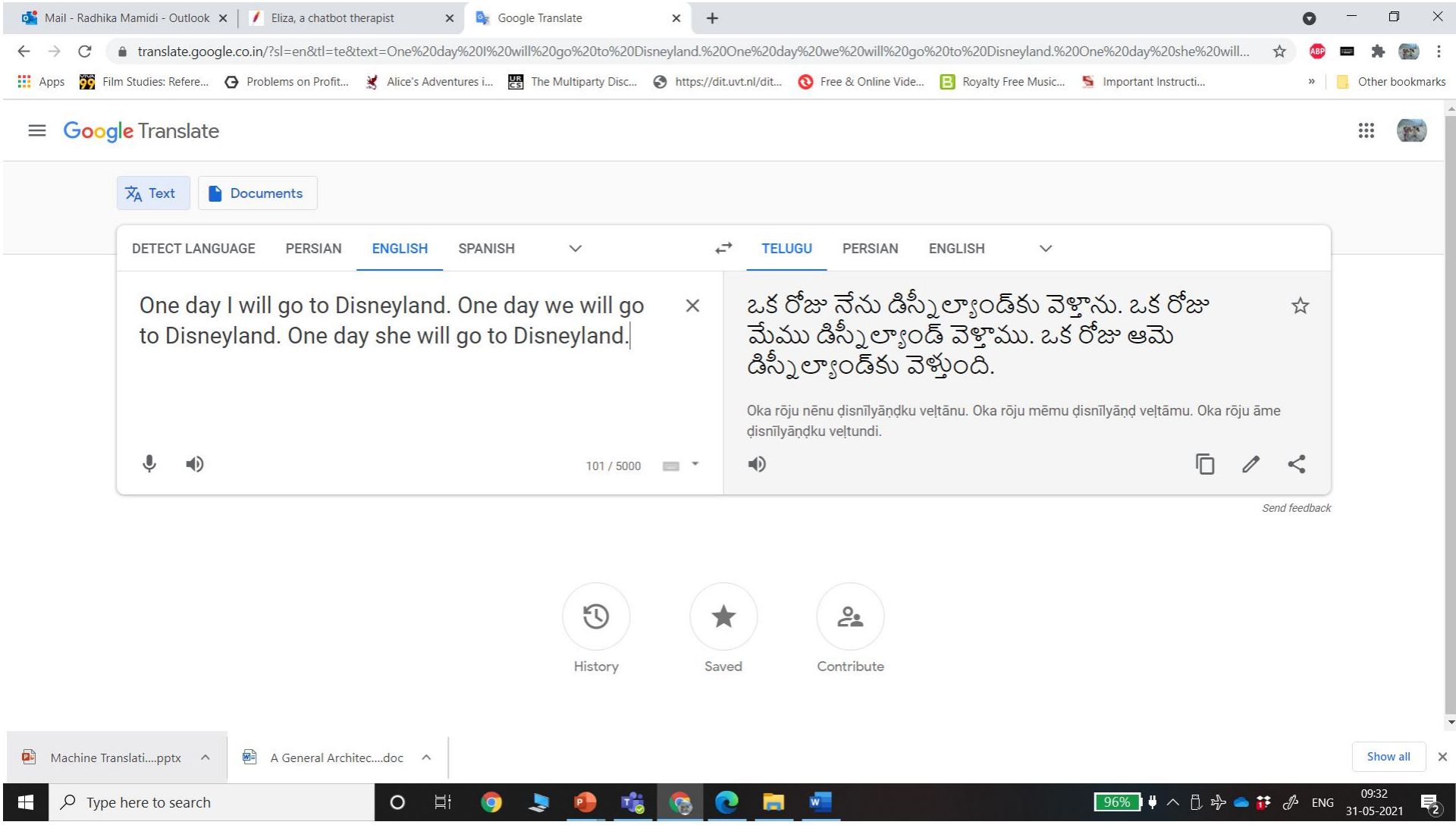
Oka rōju nēnu ḍisnīlyāṇḍ veḷatāru. Oka rōju mēm

Google Translate for Business: Translator Toolkit Website Translator Global Market Finder

Turn off instant translation

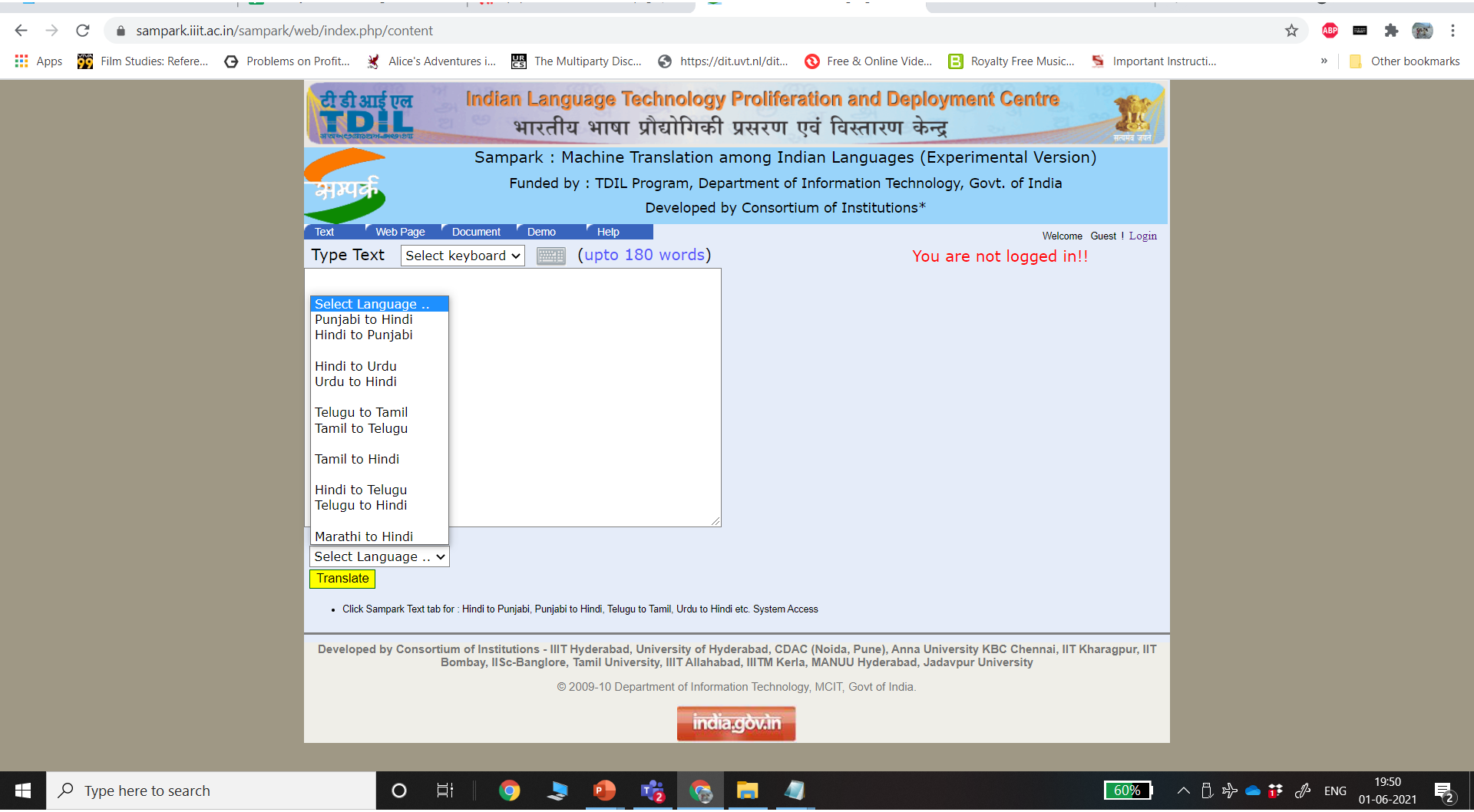
About Google Translate Mobile

Waiting for istatic.eshopcomp.com



Great difference 2014 to 2021 after incorporating Neural Network method

Sampark:
https://sampark.iiit.ac.in/sampark/web/index.php/content



MT architecture