

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.

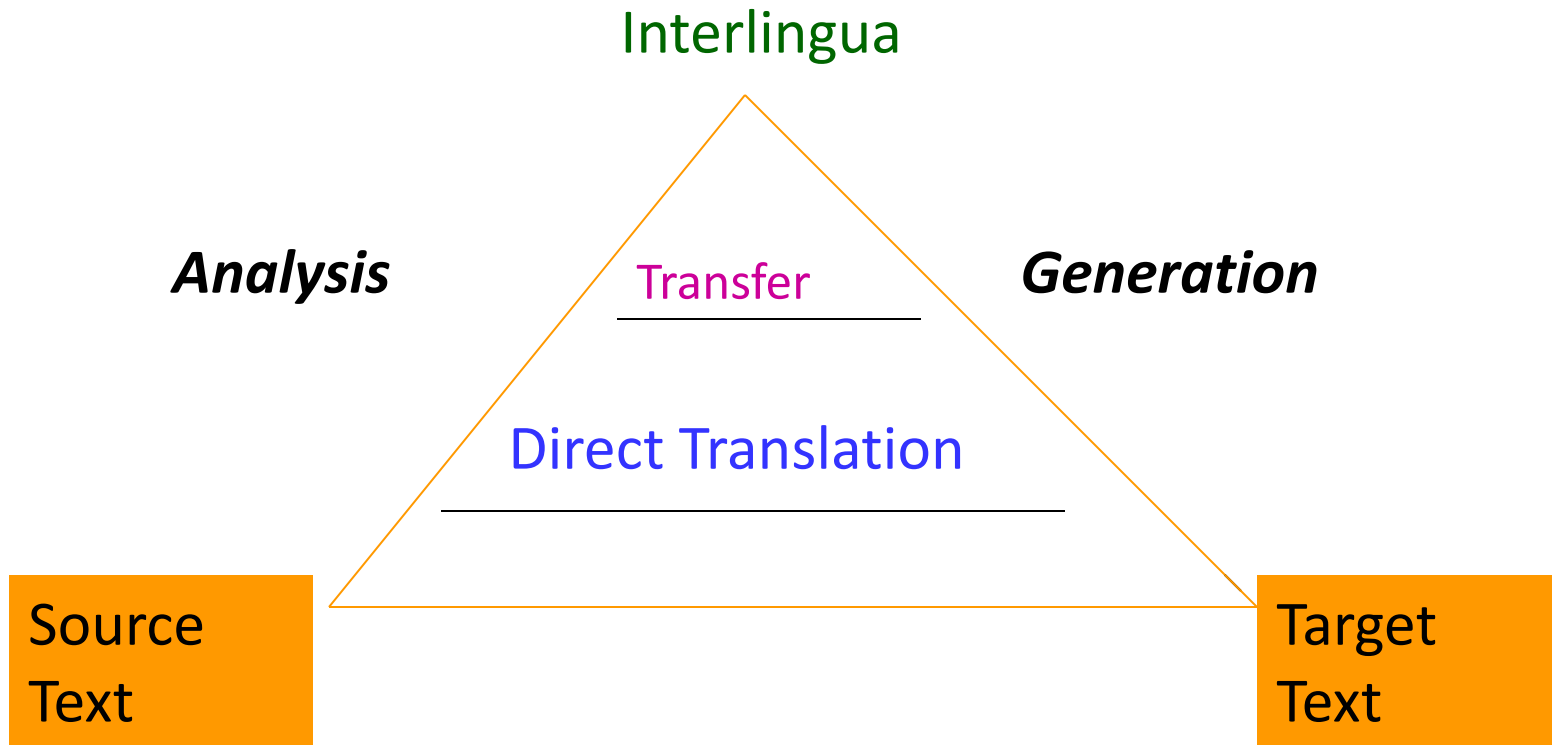
Advantages of MT

- 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



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

 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](#)

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. ✕

Type text or a website address or [translate a document](#).

ఒక రోజు నేను డిస్నీలాండ్ వెళతారు. ఒక రోజు మేము డిస్నీలాండ్ వెళతారు.



Oka rōju nēnu ḍisnīlāṇḍ vēḷatāru. Oka rōju mēmu ḍisnīlāṇḍ vēḷatāru.

The screenshot displays the Google Translate web interface. The browser's address bar shows the URL: `translate.google.co.in/?sl=en&tl=te&text=One%20day%20I%20will%20go%20to%20Disneyland.%20One%20day%20we%20will%20go%20to%20Disneyland.%20One%20day%20she%20will%20go%20to%20Disneyland.`. The Google Translate header is visible, with the 'Text' tab selected. The source language is set to 'ENGLISH' and the target language is 'TELUGU'. The input text on the left is: "One day I will go to Disneyland. One day we will go to Disneyland. One day she will go to Disneyland." The translated text on the right is: "ఒక రోజు నేను డిస్నీల్యాండ్కు వెళ్తాను. ఒక రోజు మేము డిస్నీల్యాండ్ వెళ్తాము. ఒక రోజు ఆమె డిస్నీల్యాండ్కు వెళ్తుంది." Below the translation, a phonetic transcription is provided: "Oka rōju nēnu ḍisnīlyāṇḍku vēḷtānu. Oka rōju mēmu ḍisnīlyāṇḍ vēḷtāmu. Oka rōju āme ḍisnīlyāṇḍku vēḷtundi." The interface also includes a 'Send feedback' link, a 'History' button, a 'Saved' button, and a 'Contribute' button. The Windows taskbar at the bottom shows the system clock as 09:32 on 31-05-2021, with a battery level of 96%.

Great difference 2014 to 2021 after incorporating
Neural Network method

Sampark:

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

The screenshot displays the Sampark website, an experimental machine translation tool. The header features the TDIL logo and the text "Indian Language Technology Proliferation and Deployment Centre" in English and Hindi. Below this, it states "Sampark : Machine Translation among Indian Languages (Experimental Version)" and mentions funding by the TDIL Program, Department of Information Technology, Govt. of India. The interface includes a navigation bar with tabs for Text, Web Page, Document, Demo, and Help. A "Type Text" input field is present, with a "Select keyboard" dropdown and a character count "(upto 180 words)". A "Translate" button is located below the input field. A language selection dropdown menu is open, showing options for Punjabi to Hindi, Hindi to Punjabi, Hindi to Urdu, Urdu to Hindi, Telugu to Tamil, Tamil to Telugu, Tamil to Hindi, Hindi to Telugu, Telugu to Hindi, and Marathi to Hindi. A message "You are not logged in!!" is displayed in red. The footer contains the text "Developed by Consortium of Institutions - IIIT Hyderabad, University of Hyderabad, CDAC (Noida, Pune), Anna University KBC Chennai, IIT Kharagpur, IIT Bombay, IISc-Bangalore, Tamil University, IIIT Allahabad, IIITM Kerala, MANUU Hyderabad, Jadavpur University" and "© 2009-10 Department of Information Technology, MCIT, Govt of India." The India.gov.in logo is also visible.

TDIL Indian Language Technology Proliferation and Deployment Centre
भारतीय भाषा प्रौद्योगिकी प्रसारण एवं विस्तारण केन्द्र

Sampark : Machine Translation among Indian Languages (Experimental Version)
Funded by : TDIL Program, Department of Information Technology, Govt. of India
Developed by Consortium of Institutions*

Text Web Page Document Demo Help

Welcome Guest ! Login

Type Text Select keyboard (upto 180 words)

Select Language ..
Punjabi to Hindi
Hindi to Punjabi
Hindi to Urdu
Urdu to Hindi
Telugu to Tamil
Tamil to Telugu
Tamil to Hindi
Hindi to Telugu
Telugu to Hindi
Marathi to Hindi
Select Language ..

Translate

Click Sampark Text tab for : Hindi to Punjabi, Punjabi to Hindi, Telugu to Tamil, Urdu to Hindi etc. System Access

Developed by Consortium of Institutions - IIIT Hyderabad, University of Hyderabad, CDAC (Noida, Pune), Anna University KBC Chennai, IIT Kharagpur, IIT Bombay, IISc-Bangalore, Tamil University, IIIT Allahabad, IIITM Kerala, MANUU Hyderabad, Jadavpur University

© 2009-10 Department of Information Technology, MCIT, Govt of India.

india.gov.in

MT architecture

A Simple Architecture of a Machine Translation System

