Lecture 27, 28 and 29. Machine Translation (MT)

Machine Translation (MT) refers to the **automatic translation of text or speech** from one language to another using computers. It is a core area of Natural Language Processing (NLP) and Computational Linguistics.

At its core, MT systems aim to **decode the meaning** of a sentence in a **source language (SL)** and generate its equivalent in a **target language (TL)**.

Why is Machine Translation Important?

1. Bridges language barriers in communication, helping people around the world understand each other.

2. Enables cross-cultural exchange of information (books, news, academic papers, etc.).

3. Vital for businesses, international organizations, and governments for global operations.

4. Supports real-time applications like voice assistants, email clients, and multilingual customer service.

Example of Machine Translation

Let's take a simple sentence:

- English (Source Language): *"The cat runs"*

- Spanish (Target Language): *"El gato corre"*

A machine translation system needs to:

1. Understand the words and their meanings.

2. Apply correct grammar and word order rules in the target language.

3. Output: *"El gato corre"*

Dictionary Lookup in Machine Translation

Dictionary Lookup is the most basic form of MT. It works by:

- Mapping each word in the source language to its dictionary equivalent in the target language.

- No grammar or syntax rules are considered.

Example
Given the below dictionary:

English	Spanish
the	el
a	un
cat	gato
dog	perro
runs	corre
jumps	salta
in	en
garden	jardín

Translate the following English sentences into Spanish (word-by-word).
1. *The cat runs in the garden*
2. *A dog jumps*

1. **The cat runs in the garden**
   → **el gato corre en el jardín**

2. **A dog jumps**
   → **un perro salta**

Challenges

- Each English word is matched to a Hindi word using a bilingual dictionary.

- No understanding of grammar, gender, word order, or context.

- This often leads to nonsensical or grammatically incorrect sentences.

English Word	Hindi Translation
the	वह / यह
boy	लड़का
girl	लड़की
eats	खाना (to eat)
plays	खेलना (to play)
is	है
red	लाल
car	कार
beautiful	सुंदर
apple	सेब
with	के साथ
he	वह
she	वह

Example 1: Gender Agreement Error

English Sentence:

The girl eats an apple.

Dictionary Lookup Translation:

वह लड़की खाना एक सेब ❌

Correct Hindi Translation:

वह लड़की एक सेब खाती है। ✅

Why dictionary lookup fails:

1. "खाना" is infinitive (to eat), not conjugated.

> "खाना" is infinitive (to eat), not conjugated.

This means:

- "खाना" is the **base form** of the verb — like "to eat" in English.
- It is not **conjugated**, i.e., it doesn't match the **subject**, **tense**, or **gender**.

🔄 English Parallel:

English Verb	Form Type	Example Sentence
to eat	Infinitive	I like **to eat** apples.
eats / eating / ate	Conjugated forms	She **eats** an apple.

So,

- "to eat" → **infinitive**
- "eats" → **conjugated** (present tense, third person singular)

🔁 Hindi Parallel:

Hindi Verb	Form Type	Example Sentence
खाना	Infinitive	मुझे **खाना** पसंद है।
खाता / खाती	Conjugated	वह सेब **खाती** है।

Summary:

- ✅ Use **"खाना"** when no subject agreement is needed (e.g., after "पसंद है").
- ❌ Don't use **"खाना"** as a main verb in a sentence with a subject.
- Use **खाता, खाती, खाते**, etc. depending on **gender**, **number**, and **tense**.

Since the subject is female, the verb should be खाती (not खाना).

2. Word order is wrong: S-O-V is needed in Hindi.

🔠 Word Order Comparison:

Language	Common Word Order	Meaning
English	Subject - Verb - Object (S-V-O)	E.g., "She eats an apple"
Hindi	Subject - Object - Verb (S-O-V)	E.g., "वह सेब खाती है"

Example 2: Word Order Error (SOV) agreement and
Unnecessary Article Usage

English Sentence:

The red car is beautiful.

Dictionary Lookup:

वह लाल कार है सुंदर ❌

Correct Translation:

लाल कार सुंदर है। ✅

Errors:

1. English follows **S-V-O**: *The car (S) is (V) beautiful (O),* Hindi requires **S-O-V**: *Car (S), beautiful (O), is (V)*

   But the dictionary version ends with **"सुंदर"**, which is **the object**, not the verb — this is not natural in Hindi

2. **Unnecessary Article Usage:**Insertion of "वह" (That/The):
   Hindi doesn't always need "वह" if the object is clear. So, "वह लाल कार है सुंदर" looks clunky and unidiomatic.

   Correct: लाल कार सुंदर है। (clean, natural)

Example 3: Verb Conjugation Error

English:

He plays with the boy.

Dictionary Lookup:

वह खेलना के साथ लड़का ❌

Correct Hindi:

वह लड़के के साथ खेलता है। ✅

What went wrong:

- **"खेलना" is infinitive — wrong here. It should be खेलता है.**

- **"Boy" → लड़का must become लड़के due to the postposition "के साथ".**

- **Wrong word order from literal translation.**

Example 4: Word Sense Disambiguation Failure

English:

He saw the bat.

"Bat" can mean:

- **A cricket bat (equipment)**

- **A flying mammal (जानवर)**

Dictionary lookup cannot choose between:

- बैट **(cricket bat)**

- चमगादड़ **(animal)**

So translation may wrongly become:

- उसने चमगादड़ देखा। ❌ **(meant cricket bat)**

- **Or vice versa**

Correct translation depends on context, not dictionary.

🧠 Machine Translation Challenges Summary Table (with Examples)

S.No	Challenge	Explanation	English Sentence	Incorrect Hindi Translation	Correct Hindi Translation
1.	**Literal Dictionary Lookup**	Dictionary-based MT systems translate words individually without grammar rules or context.	I want to eat food.	मैं चाहता हूं खाना खाना।	मैं खाना खाना चाहता हूं।
2.	**Word Order (S-O-V Structure)**	Hindi typically follows Subject-Object-Verb (S-O-V) structure, while English uses S-V-O.	He reads books.	वह पढ़ता किताबें है।	वह किताबें पढ़ता है।
3.	**Sentence Structure (Copula/Predicate)**	MT systems may place verbs or adjectives incorrectly.	The red car is beautiful.	वह लाल कार है सुंदर।	लाल कार सुंदर है।
4.	**Adjective-Noun Agreement**	Adjectives should match gender and number with nouns in Hindi; this is often missed.	The old man walks.	बूढ़ा आदमी चलता।	बूढ़ा आदमी चलता है।
5.	**Idioms and Figurative Language**	Literal translation of idioms leads to meaningless or funny outputs.	He kicked the bucket.	उसने बाल्टी को मारा।	उसका निधन हो गया। / वह इस दुनिया से चला गया।
6.	**Ambiguity in Word Sense**	Words with multiple meanings are mistranslated without context.	He saw a bat.	उसने एक बल्ला देखा।	उसने एक चमगादड़ देखा। (if "bat" = animal)
7.	**Cultural Nuances**	Certain expressions need culturally adapted translations to preserve intent and tone.	Let's call it a day.	आइए इसे एक दिन कहते हैं।	चलो आज का काम यहीं खत्म करें।