Why is NLP Hard?

Dr. Sambit Praharaj

Assistant Professor (II)

1. Ambiguity in Language

- Natural language has multiple interpretations.
- Lexical: 'Bank' → river bank or financial bank.
- Syntactic: 'I saw the man with the telescope.'
- Semantic: 'Every student read a book.'
- For details refer the 'Ambiguity' Slide Deck.

2. Context Dependence

- Meaning changes with context.
- 'He is cold' → temperature, emotion, personality.
- 'It is hot' → weather, food, trend.

3. Need for World Knowledge

- Understanding requires common sense.
- 'The trophy didn't fit in the suitcase because it was too big.'
- Requires reasoning about real-world sizes.

4. Sarcasm and Irony

- Machines struggle with tone and implied meaning.

- Example: 'Great! Another assignment!' (actually negative)

5. Variability in Expression

- Many ways to say the same idea.
- 'I'm hungry', 'I could eat', 'Let's grab something.'

6. Noisy, Informal Text

- Social media text is unstructured.
- 'u coming 2day bro???', mixed languages, slang.

7. Long-Range Dependencies

- Meaning depends on previous sentences.
- 'Ravi went to market. He bought vegetables.' → 'He' = Ravi

8. Figurative Language

- Metaphors, idioms, indirect speech.
- 'Break the ice', 'My phone died'.

9. Cultural Nuances

- Meaning differs by culture.
- 'That's sick!' → awesome (slang)

10. Evolving Language

- New words appear rapidly.

- 'Selfie', 'Ghosting', 'Flex'.

11. Multilingual & Code-Mixing

- Languages mixed in communication.
- 'Kal meeting fix karte hain, okay?'

12. Data Limitations

- Training data may be biased or insufficient.
- Low-resource languages suffer.