Lesson 2: Named Entity Recognition (NER) and POS Tagging

Named Entity Recognition (NER) and Part-of-Speech (POS) Tagging are essential NLP techniques for understanding the structure and semantics of language.

---

1. What is NER?

NER involves identifying named entities in text and categorizing them into predefined classes such as:

- Person (e.g., “Albert Einstein”)
- Organization (e.g., “Google”)
- Location (e.g., “Paris”)
- Date, Time, Monetary values, and more

---

2. What is POS Tagging?

- POS tagging assigns parts of speech (e.g., noun, verb, adjective) to each word in a sentence. This helps understand grammar and meaning.
- Example: Sentence: “The quick brown fox jumps over the lazy dog.” POS Tags: Det, Adj, Adj, Noun, Verb, Prep, Det, Adj, Noun

---

3. Why are these important?

- Enable information extraction and knowledge graph building
- Improve text summarization and machine translation
- Crucial for syntactic and semantic analysis

---

4. Hand-on with spaCy

- import spacy
- nlp = spacy.load("en_core_web_sm")
- doc = nlp("Apple is looking at buying U.K. startup for $1 billion")
-
- # Named Entities
- for ent in doc.ents:
- print(ent.text, ent.label_)
-
- # POS Tags
- for token in doc:
- print(token.text, token.pos_, token.tag_)

---

5. Use case in real life

- **Finance:** Extracting company names and financial terms from reports
- **Healthcare:** Identifying patient data, diseases, and medications
- **Legal:** Tagging contracts and legal clauses

---

6. Improving Accuracy

- **Use domain-specific models**
- **Fine-tune pre-trained models**
- **Combine with rule-based systems for precision**

---

7. Summary

- **NER extracts meaningful entities from text**
- **POS tagging identifies grammatical roles**
- **Both are foundational to deeper NLP tasks**