Day 11: Part-of-Speech (POS) Tagging for Urdu & Pashto

Sami Uddin Shinwari

Subject Specialist – IT | NLP Researcher

MS Data Science (FAST NU Peshawar)

BS Computer Science (UET Peshawar)

January 10, 2026

What is Part-of-Speech (POS) Tagging?

- Assigning grammatical labels to words.
- Examples: Noun, Verb, Adjective, Adverb.
- Helps machines understand sentence structure.
- A core task in Classical NLP.

Why POS Tagging is Important

- Improves text understanding.
- Used in parsing, translation, and information extraction.
- Helps disambiguate word meanings.
- Essential for downstream NLP tasks.

POS Tagging Example (English)

Sentence:
Ali reads a book

POS Tags:

- Ali – Noun
- reads – Verb
- a – Determiner
- book – Noun

POS Tagging Example (Urdu)

Sentence:

علی کتاب پڑھتا ہے

POS Tags:

- علی – (Noun)
- کتاب – (Noun)
- پڑھتا – (Verb)
- ہے – (Auxiliary Verb)

POS Tagging Example (Pashto)

Sentence:

احمد کتاب لولي

POS Tags:

- احمد – (Noun)
- کتاب – (Noun)
- لولي – (Verb)

Challenges in Urdu & Pashto POS Tagging

- Rich morphology and inflections.
- Lack of large annotated datasets.
- Free word order.
- Ambiguity due to missing diacritics.

Available Tools & Resources

- CRULP Urdu POS Tagger
- IIIT Hyderabad Urdu NLP tools
- Universal Dependencies (limited support)
- Rule-based and ML-based taggers

Why POS Tagging is Harder for Low-Resource Languages

- Few labeled corpora.
- Limited pretrained models.
- Dialectal variations.
- Lack of standardization.

Key Takeaways

- POS tagging is a backbone NLP task.
- Urdu & Pashto need more annotated datasets.
- Classical models still matter.
- Community-driven resources are essential.

#Day11 #POSTagging #UrduNLP #PashtoNLP #LowResourceNLP