Morphology, Stemming & Lemmatization

Dr. Sambit Praharaj

Assistant Professor (II)

What is Morphology?

- Morphology studies word structure
- Words are made of morphemes
- Morphemes carry meaning or grammar

- Example:
- walk + ed → walked

Types of Morphemes

- Free morphemes: walk, play
- Bound morphemes: -ed, -ing, -s

- Example:
- play + ing → playing

Bound: Inflectional Morphology

- Adds grammatical meaning
- Tense, number, comparison

- Examples:
- play → played
- cat → cats
- big → bigger

Bound: Derivational Morphology

- Creates new words or changes Part Of Speech (POS)

- Examples:
- happy → happiness
- teach → teacher
- nation → national

Why Morphology Matters in NLP?

- Reduces word variation
- Preserves meaning
- Essential for normalization
- Critical for rich-morphology languages

What is Stemming?

- Rule-based suffix stripping
- No linguistic understanding
- Fast but crude

Stemming Examples (Porter)

- studies → studi
- studying → studi
- university → univers
- universal → univers

Problems with Stemming

- Produces non-words
- Ignores meaning
- No POS awareness

Over-stemming Example

- compute
- computer
- computation

- All reduced to: comput
- (Meaning is lost)

Under-stemming Example

- go → go
- went → went

- Related words not merged

Why Stemming Fails Without Morphology?

- Cannot distinguish inflection vs derivation
- Removes meaningful parts of words

What is Lemmatization?

- • Finds dictionary base form (lemma)
- • Uses morphology + POS
- • Linguistically accurate

Lemmatization Examples

- running → run
- ran → run
- mice → mouse
- better → good

Why Lemmatization Needs Morphology?

- Handles irregular forms (go → went)
- Preserves meaning
- Requires word knowledge

Stemming vs Lemmatization

- Stemming:
  - Fast
  - Inaccurate
  - Non-words
- Lemmatization:
  - Accurate
  - Slower
  - Real words

Inflection vs Derivation

- Inflection (safe to normalize):
- played, playing

- Derivation (meaning change):
- happiness, teacher

Impact on NLP Applications

- Search engines
- Machine translation
- POS tagging
- Language Modeling
- Information retrieval

Final Takeaway

- Morphology tells us:
- What can be removed safely
- What must be preserved

- Without morphology, normalization is guesswork