In **NLP (Natural Language Processing)**, **morphology** is the study of the **structure and formation of words**.

It focuses on how **words are built** from smaller units called **morphemes**, where:

* A **morpheme** is the smallest meaningful unit in a language.
  + Example:
    - "unhappiness" → made of three morphemes:  
      **un-** (prefix meaning "not") + **happy** (root word) + **-ness** (suffix turning adjective to noun).

| **Type** | **Description** | **Example** |
| --- | --- | --- |
| **Inflectional morphology** | Modifies a word’s tense, number, aspect, etc., but keeps the core meaning. | play → played, dog → dogs |
| **Derivational morphology** | Forms a new word by adding affixes, often changing the word’s category or meaning. | happy → happiness, teach → teacher |

**1. Stemming**

* **Meaning**:  
  Stemming **chops off** the ends of words to find the **root form**, but it doesn't always produce a real dictionary word.
* **Goal**:  
  Quickly reduce a word to its **stem** (basic form).
* **How**:  
  It uses simple rules like removing "ing", "ed", "s", etc.

**Example**:

* "playing" → "play"
* "played" → "play"
* "happiness" → "happi" (⚡ not a real word!)

**Note**:  
Stemming can be a bit rough — sometimes it cuts too much or too little.

**2. Lemmatization**

* **Meaning**:  
  Lemmatization **reduces words** to their **base/dictionary form** (**lemma**), and the result is a **real word**.
* **Goal**:  
  Get a **proper root** considering the **word’s meaning and grammar** (like whether it's a noun, verb, etc.).
* **How**:  
  It uses **vocabulary** and **morphological analysis** of the word.

**Example**:

* "playing" → "play"
* "better" → "good"
* "studies" → "study"

**What is Punkt?**

* **Punkt** is a **pre-trained model** for **tokenization** in **NLTK** (Natural Language Toolkit) library.
* It is mainly used to **split text into sentences** and **words**.
* It **learns** how sentences are divided based on language data — without relying only on rules (like always splitting at every period).

**What are POS tags?**

* **POS** stands for **Part of Speech**.
* **POS tagging** means **labeling each word** in a sentence with its **grammatical role** (like noun, verb, adjective, etc.).