**Batch: H-ADS (H2-3)**

**Roll No.: 16010122221**

**Experiment No. 10**

**Title : Virtual lab on text processing (word analysis)**

**Aim:** To implement text processing (Word analysis)

**Expected Outcome of Experiment:**

**CO4 :** Understand the basic concept and techniques of Machine Learning clustering.

**Books/ Journals/ Websites referred:**

1. https://nlp-iiith.vlabs.ac.in/exp/word-analysis/index.html

**Introduction**

A word can be simple or complex. For example, the word 'cat' is simple because one cannot further decompose the word into smaller part. On the other hand, the word 'cats' is complex, because the word is made up of two parts: root 'cat' and plural suffix '-s'



**Procedure**

STEP 1: Select the language.

OUTPUT: Drop down for selecting words will appear. STEP 2: Select the word.

OUTPUT: Drop down for selecting features will appear. STEP 3: Select the features.

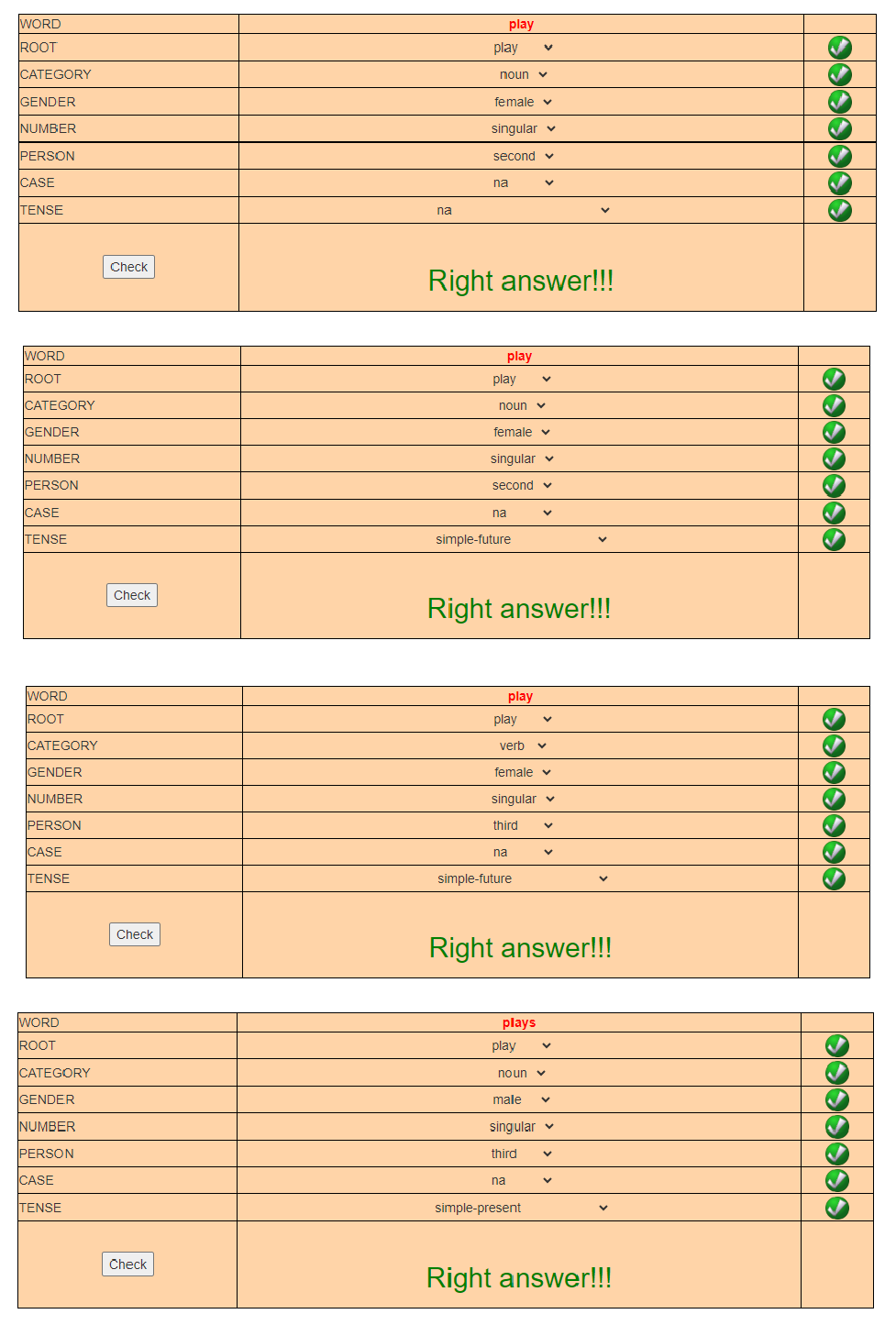
STEP 4: Click "Check" button to check your answer.

OUTPUT: Right features are marked by tick and wrong features are marked by cross.

**Simulation:**

**English:**





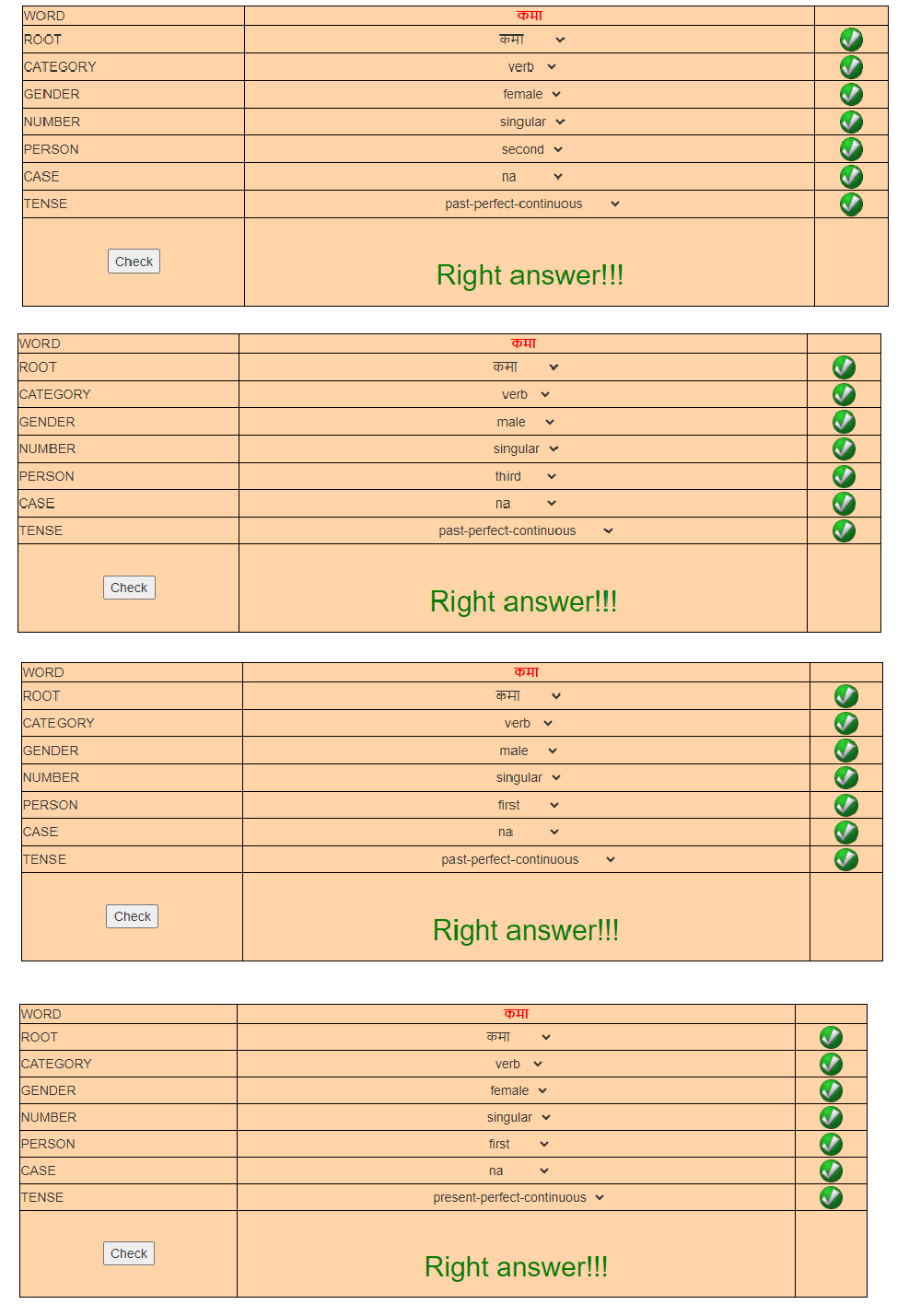






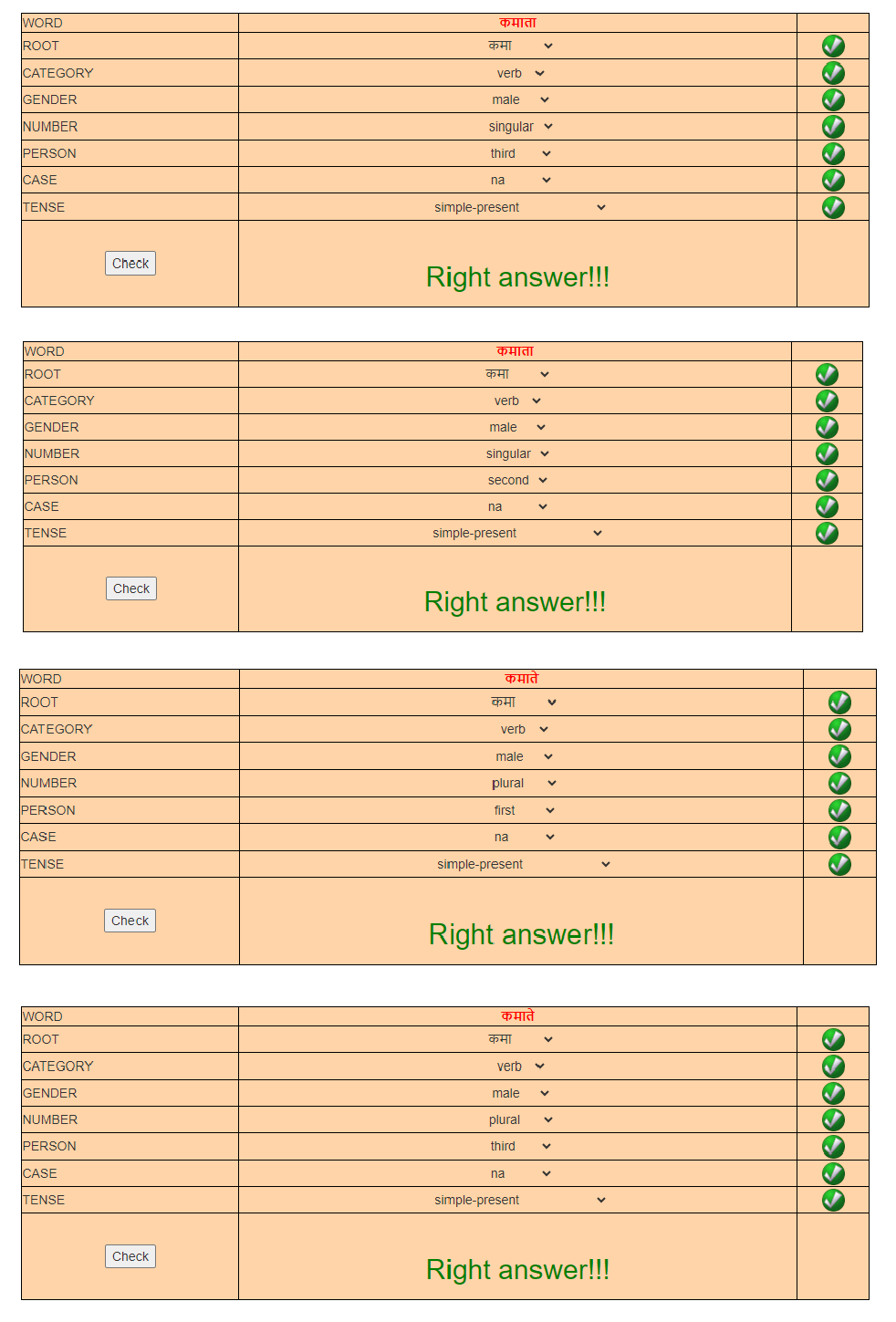


**Hindi:**

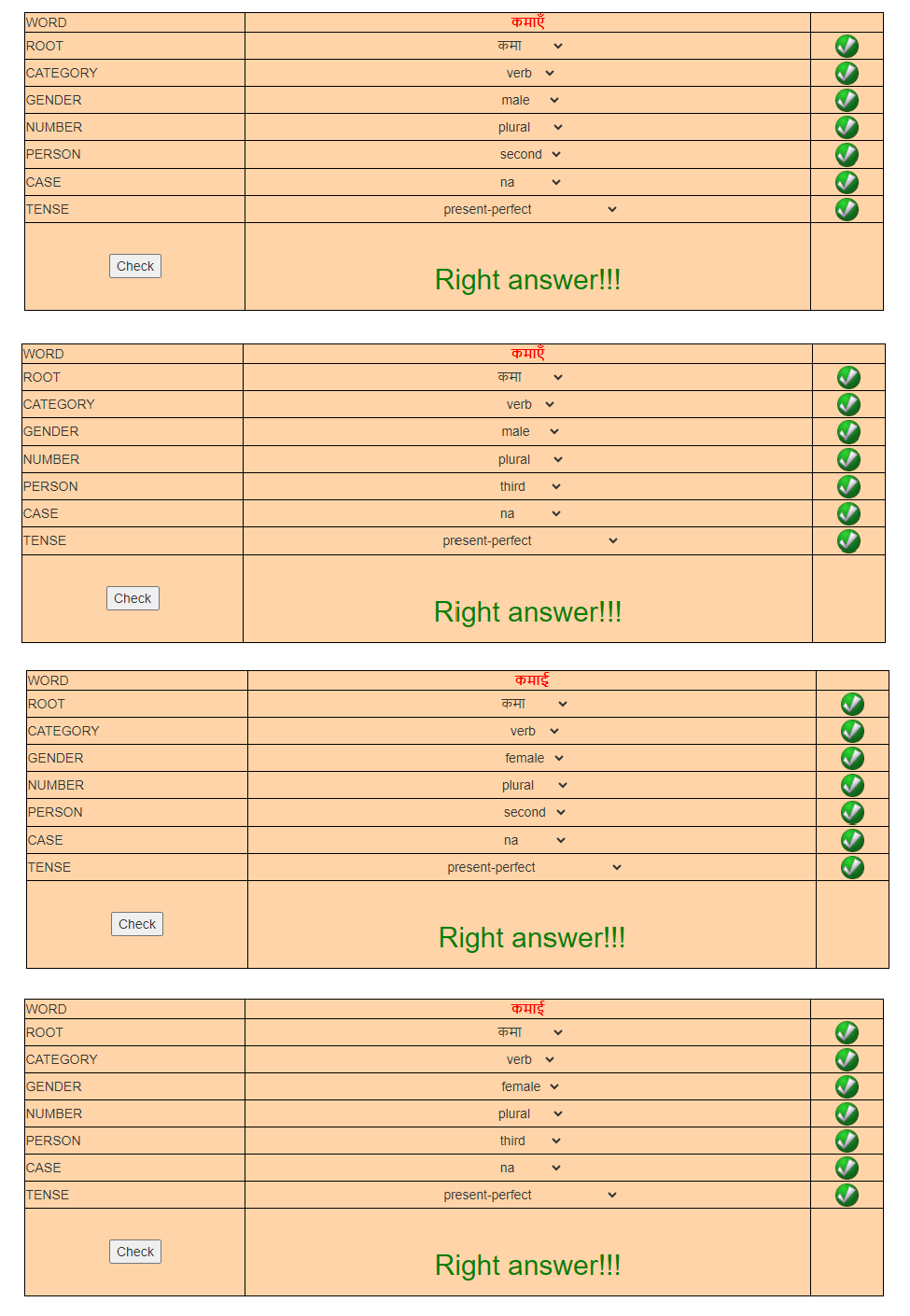
















**Assignment:**

# Choose a typical masculine noun, ending in 'A', from your language. Write down its various forms along with various features and their values associated with them.

**Answer:**

# In Gujarati, a typical masculine noun ending in 'a' could be "વિદ્યા" (vidyā), which means "knowledge" in English. Here are its various forms along with associated features and their values:

Nominative Singular: વિદ્યા (vidyā) Gender: Masculine

Number: Singular

Case: Nominative (subject case) Definiteness: Indefinite

Nominative Plural: વિદ્યાઓ (vidyāo) Gender: Masculine

Number: Plural

Case: Nominative (subject case) Definiteness: Indefinite

Accusative Singular: વિદ્યા (vidyā) Gender: Masculine

Number: Singular

Case: Accusative (direct object case) Definiteness: Indefinite

Accusative Plural: વિદ્યાઓ (vidyāo) Gender: Masculine

Number: Plural

Case: Accusative (direct object case) Definiteness: Indefinite

Genitive Singular: વિદ્યાનો (vidyāno) Gender: Masculine

Number: Singular

Case: Genitive (possessive case) Definiteness: Indefinite

Genitive Plural: વિદ્યાઓનો (vidyāono) Gender: Masculine

Number: Plural

Case: Genitive (possessive case)

Definiteness: Indefinite

Dative Singular: વિદ્યાએ (vidyāe) Gender: Masculine

Number: Singular

Case: Dative (indirect object case) Definiteness: Indefinite

Dative Plural: વિદ્યાઓએ (vidyāoae) Gender: Masculine

Number: Plural

Case: Dative (indirect object case) Definiteness: Indefinite

These forms cover the various grammatical cases and numbers that the noun "વિદ્યા" (vidyā) can take in Gujarati.

# English has a suffix -en whose use is illustrated in the following lists:

|  |  |
| --- | --- |
| **List A** | **List B** |
| **Red** | **Redden** |
| **Mad** | **Madden** |
| **Soft** | **Soften** |
| **Wide** | **Widen** |
| **Sharp** | **Sharpen** |

**In regard to these data, answer the following questions:**

# What part of speech does the suffix -en attach to? That is, what is the part of speech of the words in list A?

The suffix "-en" typically attaches to adjectives or descriptive words to form verbs. In the given lists:

List A contains adjectives or descriptive words such as "Red," "Mad," "Soft," "Wide," and "Sharp."

List B contains verbs formed by adding the suffix "-en" to the adjectives in List A: "Redden," "Madden," "Soften," "Widen," and "Sharpen."

Therefore, the part of speech of the words in List A is adjectives, and the suffix "-en" attaches to them to form verbs in List B.

# When the suffix -en is attached to a word, what part of speech is the resulting word? Give some specific morphological properties of one of the words in list B, in order to justify your answer.

When the suffix "-en" is attached to a word, the resulting word typically becomes a verb.

For example, let's consider the word "Sharpen" from List B:

Part of Speech: The resulting word "Sharpen" is a verb. Morphological Properties:

Base Word: The base word is "Sharp," which is an adjective describing the quality of being sharp.

Suffix: The suffix "-en" has been added to the base word "Sharp" to form the verb "Sharpen."

Meaning: The addition of "-en" changes the meaning of the word from describing a quality (sharp) to an action (to make something sharp).

Inflection: The verb "Sharpen" can be inflected for tense, aspect, and mood, such as "Sharpened" (past tense), "Sharpening" (present participle), and "Will sharpen" (future tense).

Derivational Morphology: The addition of the suffix "-en" creates a derived word (verb) from the base adjective "Sharp."

Pronunciation: The pronunciation of the word changes slightly due to the addition of the suffix, with stress often shifting to the first syllable ("SHAR-pen" instead of "sharp").

These morphological properties demonstrate that "Sharpen," and similarly the other words in List B formed by adding "-en" to adjectives in List A, are verbs.

# Take one verb from your mother tongue, gloss it (i.e., give the Engish meaning) and conjugate it in all tenses and aspects and persons.

Let's take the verb "खाना" (khaana) from Hindi. The gloss or English meaning of "खाना" is "to eat."

Now, let's conjugate it in all tenses, aspects, and persons:

Present Indefinite:

I eat: मैं खाता हूँ (main khaata hoon) [masculine], मैं खाती हूँ (main khaati hoon) [feminine]

You eat: तुम खाते हो (tum khaate ho) [informal], आप खाते हैं (aap khaate hain) [formal]

He/She/It eats: वह खाता है (vah khaata hai) [masculine], वह खाती है (vah khaati hai) [feminine]

Present Continuous:

I am eating: मैं खा रहा हूँ (main kha raha hoon) [masculine], मैं खा रही हूँ (main kha rahi hoon) [feminine]

You are eating: तुम खा रहे हो (tum kha rahe ho) [informal], आप खा रहे हैं (aap kha rahe hain) [formal]

He/She/It is eating: वह खा रहा है (vah kha raha hai) [masculine], वह खा रही है (vah kha rahi hai) [feminine]

Present Perfect:

I have eaten: मैंने खा लिया है (maine kha liya hai)

You have eaten: तुमने खा लिया है (tumne kha liya hai) He/She/It has eaten: वहने खा लिया है (vahne kha liya hai)

Past Indefinite:

I ate: मैंने खाया (maine khaaya)

You ate: तुमने खाया (tumne khaaya) He/She/It ate: उसने खाया (usne khaaya)

Past Continuous:

I was eating: मैं खा रहा था (main kha raha tha) [masculine], मैं खा रही थी (main kha rahi thi) [feminine]

You were eating: तुम खा रहे थे (tum kha rahe the) [informal], आप खा रहे थे (aap kha rahe the) [formal]

He/She/It was eating: वह खा रहा था (vah kha raha tha) [masculine], वह खा रही थी (vah kha rahi thi) [feminine]

Future Indefinite:

I will eat: मैं खाऊूँ गा (main khaaunga) [masculine], मैं खाऊूँ गी (main khaaungi) [feminine] You will eat: तुम खाओगे (tum khaaoge) [informal], आप खाएं गे (aap khaayenge) [formal] He/She/It will eat: वह खायेगा (vah khaayega) [masculine], वह खायेगी (vah khaayegi) [feminine]

These are some conjugations of the verb "खाना" (khaana) in various tenses, aspects, and persons in Hindi.

# Refer to the following data and answer the question below:

**List 1: taller, shorter, higher, lower, smarter <br> List 2: mower, teacher, sailor, caller, operator <br> List 3: never, cover, finger, river <br>**

# Are the words ending with 'er'/'or' have some common features?

Yes, the words ending with 'er'/'or' in the given lists share a common feature: they are typically agent nouns or comparative adjectives.

Agent Nouns: Words ending in '-er' or '-or' often denote a person or thing that performs a particular action or is associated with a particular role or profession.

For example:

"mower" (someone who mows grass) "teacher" (someone who teaches) "sailor" (someone who sails)

"caller" (someone who calls) "operator" (someone who operates)

Comparative Adjectives: In the context of comparative adjectives, words ending in '- er' or '-or' indicate a comparison between two things or qualities.

For example:

"taller" (comparative form of 'tall') "shorter" (comparative form of 'short') "higher" (comparative form of 'high') "lower" (comparative form of 'low') "smarter" (comparative form of 'smart')

These endings often denote the person or thing performing an action (agent nouns) or indicate a comparison between two qualities (comparative adjectives).

# Identify root and suffix in the following words: kissed

**stronger goodness teacher achievement**

Here are the root and suffix for each word: kissed:

Root: kiss Suffix: ed

stronger:

Root: strong Suffix: er

goodness:

Root: good Suffix: ness

teacher:

Root: teach Suffix: er

achievement:

Root: achieve Suffix: ment

These are the root and suffix for each word provided**.**

**Post Lab Questions:**

1. **List any two real-time applications of NLP?**

Two real-time applications of Natural Language Processing (NLP) are:

Chatbots and Virtual Assistants: NLP is extensively used in creating chatbots and virtual assistants that can understand and respond to natural language inputs. These applications are used in various industries for customer service, technical support, sales, and information retrieval. Examples include virtual assistants like Siri, Alexa, and Google Assistant, as well as chatbots on websites and messaging platforms.

Sentiment Analysis: NLP is used for sentiment analysis to analyze and understand the sentiment or opinion expressed in text data, such as social media posts, product reviews, and customer feedback. Sentiment analysis tools can automatically classify text as positive, negative, or neutral, enabling businesses

to gauge public opinion, monitor brand perception, and make data-driven decisions. This application is widely used in market research, social media monitoring, customer experience management, and reputation management.

# What are stop words?

Stop words are commonly used words in a language that are typically filtered out or excluded from text processing and analysis because they do not carry significant meaning or contribute much to the overall understanding of the text. These words are extremely common and appear frequently across different types of text documents, regardless of the topic or context. Examples of stop words in English include "the," "is," "and," "of," "in," "to," "a," etc.

In Natural Language Processing (NLP), stop words are often removed from text data during preprocessing to improve the efficiency of algorithms and to focus on the more important words that carry semantic meaning. Removing stop words can also reduce the dimensionality of the data and improve the performance of tasks such as text classification, information retrieval, and sentiment analysis. However, the list of stop words can vary depending on the specific application or domain.

# List the components of Natural Language Processing.

The components of Natural Language Processing (NLP) typically include:

Text Preprocessing: This involves cleaning and preparing raw text data for further analysis. It may include tasks such as tokenization (breaking text into words or sentences), lowercasing, removing punctuation, and removing stop words.

Lexical Analysis: Also known as tokenization, this step involves breaking down the text into smaller units such as words, phrases, or symbols, called tokens. This process helps in understanding the structure of the text and extracting meaningful information.

Morphological Analysis: This component deals with analyzing the structure and form of words, including their root, prefixes, suffixes, and grammatical variants. Stemming and lemmatization are common techniques used in morphological analysis to normalize words to their base or dictionary form.

Syntactic Analysis (Parsing): This involves analyzing the grammatical structure of sentences to determine how words relate to each other and form meaningful phrases and sentences. Parsing helps in understanding the syntax and semantics of the text.

Semantic Analysis: This component focuses on understanding the meaning of words and sentences in context. It involves tasks such as named entity recognition (identifying named entities like people, organizations, and locations), word sense disambiguation, and semantic role labeling.

Discourse Analysis: This involves analyzing the structure and flow of text beyond the sentence level, including how sentences are connected and organized to convey coherent meaning. Discourse analysis helps in understanding the overall context and intention of the text.

Pragmatic Analysis: This component deals with analyzing the meaning of text in relation to the context, background knowledge, and communicative goals. It involves understanding implied meaning, sarcasm, ambiguity, and other aspects of language use that go beyond literal interpretation.

Text Generation: This involves generating human-like text based on input data or predefined patterns. Text generation can be used for tasks such as machine translation, summarization, question answering, and dialogue systems.

These components are often interconnected and may involve the use of various machine learning and deep learning techniques to analyze and process natural language data effectively.

**Conclusion:**

The implementation of text processing for word analysis proved to be instrumental in uncovering valuable insights from raw textual data. Through techniques such as tokenization, lexical analysis, and morphological analysis, we were able to break down and understand the structure of words within the text. This facilitated tasks such as sentiment analysis, named entity recognition, and syntactic parsing, empowering us to extract meaningful information and enhance our understanding of the underlying semantics within the text. Overall, the experiment demonstrated the effectiveness of text processing techniques in unlocking the potential of natural language data for various applications across domains.