```
#excller day-8
# Install and import NLTK
!pip install nltk
import nltk

# Download the 'punkt_tab' tokenizer models from NLTK
nltk.download('punkt_tab') # Changed from 'punkt' to 'punkt_tab'

# Function to tokenize text into sentences and words
def tokenize_text(paragraph):
    # Tokenize paragraph into sentences
    sentences = nltk.sent_tokenize(paragraph)
    # Tokenize each sentence into words
    words = [nltk.word_tokenize(sentence) for sentence in sentences]
    return sentences, words

# Input paragraph
paragraph = """
Tokenization is the process of breaking text into smaller units called tokens.
These tokens can be words, sentences, or even smaller units. Tokenization is an
important step in text preprocessing.
"""

# Tokenize the paragraph
sentences, words = tokenize_text(paragraph)

# Print the tokenized sentences
print("Sentences:")
for sentence in sentences:
    print(sentence)

# Print the tokenized words for each sentence
print("\nWords:")
for word_list in words:
    print(word_list)
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages (3.9.1)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk) (8.1.8)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk) (4.67.1)
[nltk_data] Downloading package punkt_tab to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt_tab.zip.
Sentences:

Tokenization is the process of breaking text into smaller units called tokens.
These tokens can be words, sentences, or even smaller units.
Tokenization is an
important step in text preprocessing.

Words:
['Tokenization', 'is', 'the', 'process', 'of', 'breaking', 'text', 'into', 'smaller', 'units', 'called', 'tokens', '.']
['These', 'tokens', 'can', 'be', 'words', ',', 'sentences', ',', 'or', 'even', 'smaller', 'units', '.']
['Tokenization', 'is', 'an', 'important', 'step', 'in', 'text', 'preprocessing', '.']
```

Start coding or generate with AI.