```
import nltk
nltk.download("punkt")
nltk.download("stopwords")
nltk.download("wordnet")
nltk.download("averaged_perceptron_tagger")

text = "Tokenization is the first step in text analytics. The process of breaking down a text paragraph into smaller chunks such as words or sentences is called Tokenization."

#Sentence Tokenization
from nltk.tokenize import sent_tokenize

tokenized_text = sent_tokenize(text)
print(tokenized_text)

#Word Tokenization
from nltk.tokenize import word_tokenize

tokenized_word = word_tokenize(text)
print(tokenized_word)

#print stop words of English
from nltk.corpus import stopwords
stop_words = set(stopwords.words("english"))
print(stop_words)

import re
text = "How to remove stop words with NLTK library in Python?"

text = re.sub('[^a-zA-Z]', ' ', text)
print(text)
```

```
tokens = word_tokenize(text.lower())
filtered_text = []
for w in tokens:
  if w not in stop_words:
    filtered_text.append(w)
print("tokenized Sentence:", tokens)
print("Filtered Sentence:", filtered_text)

#Stemming
from nltk.stem import PorterStemmer
e_words = ["wait", "waiting", "waited", "waits"]
ps = PorterStemmer()
for w in e_words:
  rootWord = ps.stem(w)
  print(rootWord)

#Lemmatization
from nltk.stem import WordNetLemmatizer
wordnet_lemmatizer = WordNetLemmatizer()
text = "studies studying cries cry"
tokenization = nltk.word_tokenize(text)
for w in tokenization:
  print("Lemma for {} is {}".format(w, wordnet_lemmatizer.lemmatize(w)))
```