

Text processing in Natural Language Processing (NLP) involves a series of techniques and methods used to prepare and manipulate text data for various NLP tasks like language translation, sentiment analysis, text summarization, and more.

1. **Tokenization:** This is the process of breaking down text into smaller units called tokens. Tokens can be words, characters, or subwords. This step is fundamental for most NLP tasks as it converts unstructured text into a structured form.
2. **Normalization:** This step involves converting text into a more uniform format. It includes processes like converting all characters to lower or upper case, removing punctuation, or converting numbers to words. This helps in reducing the complexity of the text data.
3. **Stop Words Removal:** Stop words are common words like "is", "and", "the", etc., that are often removed from the text as they usually don't contribute much to the meaning of a sentence for many NLP tasks.
4. **Stemming and Lemmatization:** Both are techniques to reduce words to their base or root form. Stemming crudely chops off prefixes and suffixes (e.g., "running" to "run"). Lemmatization, on the other hand, involves a more sophisticated linguistic approach to convert a word to its base or dictionary form (e.g., "better" to "good").
5. **Part-of-Speech Tagging:** This is the process of identifying parts of speech (like nouns, verbs, adjectives) in the text. It's useful in many NLP applications such as parsing and word sense disambiguation.
6. **Named Entity Recognition (NER):** NER involves identifying and categorizing key information (names of people, organizations, locations, etc.) in text. This is particularly useful in information extraction tasks.
7. **Syntactic Parsing:** This involves analyzing the grammatical structure of a sentence, identifying relationships between words, and constructing a parse tree. This is crucial for understanding the syntactic structure of sentences.
8. **Semantic Analysis:** This step involves understanding the meaning and interpretation of words in context. It includes tasks like word sense disambiguation and semantic role labeling.
9. **Feature Extraction:** This involves converting text into a set of features (numerical or categorical) for use in modeling. Techniques include Bag-of-Words, TF-IDF (Term Frequency-Inverse Document Frequency), and word embeddings like Word2Vec or GloVe.
10. **Word Embeddings:** These are vector representations of words that capture contextual meanings, syntactic and semantic relationships. Models like Word2Vec, GloVe, and BERT provide pre-trained embeddings which can be used for various NLP tasks.

**11. Text Classification and Clustering:** Classification involves categorizing text into predefined classes, while clustering involves grouping similar texts together. These are common tasks in NLP for applications like spam detection, sentiment analysis, and topic modeling.