1.自然語言處理的一些典型問題

（1）容錯力

打錯字

語法錯誤

未知詞（沒有被收錄的詞彙）

（2）強健性

領域改變

網絡語言：方言俗語、各門外語、縮略語、諧音、符號合併

表情符號

2.語言單位、成分、或單元

詞素（Morpheme）

字元（Character）

詞彙（Word）

n-連詞（N-gram）

多詞表達（Multiple Word Expression）

命名實體（Named Entity）

子句（Clause）

句子（Sentence）

段落（Paragraph）

段落（Passage）

文件（Document）

多文件（Multi-Document）

3.分類：掌握共通性

詞性類別

語義類別

句法類別

相依類別

言談類別

意見類別

情感類別

立場類別

4.自然語言處理一些任務

詞素（Morpheme）->詞素分析（Morpheme Analyzer）

詞彙（Word）->中文斷詞（Tokenizer/Chinese Segmentation）

命名實體（Named Entity）->命名實體辨識（Named Entity Recognition）

詞性（Part of Speech）->詞性標記（Part of Speech Tagger）

句法類別（Syntactic Category）->句法剖析（Syntactic Category/Chunker）

相依類別（Dependency Category）->相依剖析（Dependency Parser）

語義類別（Semantic Category）->語義關係標記（Semantic Role Labelling）

言談關係（Discourse Relation）->言談剖析（Discourse Parser）

意見（Opinion）->意見勘探（Opinion Mining）

情感（Emotion）->情感分析（Emotion Analysis）

立場（Stance）->立場偵測（Stance Detector）

5.符號計算

詞彙和類別都是以符號呈現

匹配是基本符號計算

符號匹配無法計算詞彙語義關聯程度（semantic relatedness）

6.分佈式表達（Distributional Representation）

意思（meanings）

意思的產生來自使用！（Meaning as use!）

要了解詞彙的意思，關鍵是伴隨出現的詞彙（You shall know a word by the company it keeps.(J.R.Firth, 1957)）

意思相似的詞彙很可能出現在相似的語境（words with similar meanings are likely to appear in similar contexts）

分佈式假設（Distributional hypothesis）:如果兩個詞的上下文相似，則這兩個詞的詞義是相似。

語境（上下文）

7.分散式表示（Distributed Representation）

向量空間的詞彙表示

Continuous Bag of Words(CBOW)

用上下文的詞來預測目標詞

Skip-gram

以當前的詞來預測上下文的詞

將詞彙轉換成低緯度稠密向量（word2vec）