# 1 Semantic Parsing

**Note: This section is fairly rough. It may or may not be updated in future**

**Natural language is ambiguous, ambiguity makes computational methods difficult. Therefore a traditional approach for semantic parsing is to first map the natural language to some unambiguous logical statement, e.g. prepositional logic or first order logic.**

## 1.1 Semantic Role Labelling - Shallow Semantic Parsing

**SRL is the process of labelling each noun phrase that is an argument to the verb (basically all of them). Some typical roles include:**

- agent
- patient
- source
- destination
- instrument

**SRL can be useful in assisting question answering, information retrieval, text summarisation, machine translation...**

### 1.1.1 Approaches

1. Hand coded rules based on grammatical knowledge, e.g. (Syntactic) the Agent is often the subject, (Sectional restrictions / Verb constraints) Agents should be animate.. etc
2. Statistical methods as a sequence labelling task (HMM / CRF / Token classifiers)
3. (If a parse tree is available) using a parse tree as features with statistical methods

## 1.2 Semantic Parsing - Proper

**The task of transforming natural language into completely formal meaning representations**

#### 1.2.1 Example Formats

- **Predicate Logic Query Language**
  "What is the smallest state by area?"
  $answer(x_1, smallest(x_2, state(x_1), area(x_1, x_2)))$
- **Functional Query Language (functional and variable free)**
  "What is the smallest state by area?"
  $answer(smallest\_one(area(state(all)))$

#### 1.2.2 Approaches

1. Supervised training with annotated pairs
2. Compositional methods building an MR recursively bottom up using a parse tree
3. As a word alignment/machine translation problem (WASP) using e.g. IBM model 5
4. Using CFSG's (Context Free Semantic Grammars) with CKY algorithm:
   (a) Annotate the data set as necessary, grouping by category (E.g. "I": Person, "like": verb, "milk": food)
   (b) Define ALL possible rules from the annotations and sentences only ($\$Person \rightarrow I, Mel$)
   (c) Break into CNF form as needed using placeholder variables
   (d) Use CKY to build up a semantic tree.

# 2 Lexical Semantics

## 2.1 Word Sense Disambiguation

#### 2.1.1 Learning for WSD

1. Treat as a (unsupervised) classification problem, assume that P.O.S. tags have been determined for target word
2. Extract useful features for the classifier:
   - Surrounding bag of words (BOW)
   - P.O.S. of neighbouring words
   - Local co-locations
   - Syntactic relations
3. Annotate a small set of words by hand, and use this to train a semi-supervised classifier using the Yarowski algorithm with expectation maximisation

## 2.2 Coreference Resolution

**Coreference resolutoin is the task of finding all expressions that refer to the same entity in a body of text.**

### 2.2.1 High level approach

1. Detect "mentions" (not too difficult). Use POS tags, NER extraction, parse trees and cast a wide net
2. Cluster the mentions (tricky) using either a labelled corpus and supervised classifier ("Mention Ranking"), or an unsupervised clustering approach

### 2.2.2 Mention Pairs and Mention Ranking

**Given a mention, we would like to find the most likely entity that it refers to. We assume that there is a labelled corpus (although it is possible to use an unlabelled corpus with EM and counts). To simplify the search space, we apply some simple heuristics:**

- Recency (more likely to refer to a recent entity)
- Centering (more likely to refer to the central document entity)
- Gender/number agreement
- Grammatical role
- Parallelism