* Representational system: The sense through which we expirence the world are refered to as representational system.

* The words used to describe these system are called predicate.

(OR)

* predicates are the words that used to describes our senses like

1. Visual (what we see)
2. Additory (which we hear)
3. Kinaesthetic (which we can feel)
4. Self-talk (Labelling system)
5. Olfactory (things we smell)
6. gustatory (what we taste)

* these are also called as process words or sense specific words.

* visual predicates like - see, view, clear, illeminate, shawed, appeard etc..

* kinaesthetic predicate like - felt, grasp, hard, rough, shape, turnover.

* Auditory predicate like - heard, silence, listens, tunes, resonates etc.

* Olfactory predicate like - smell, fragrance, odour, sniff etc.

*Self-talk predicate like or (Auditory digital predicate) like think, decide, motivate, consider, understand, know etc..

Argument: An argument is an expression that helps in completing the meaning of a predicate.

* A predicate and its argument form a predicate argument structure.

for example: Rohan (S) like (O) oranges here two arguments are there that subject argument and object argument.

Predicate Argument Structure:

1. Frame Net: It produces on electical (resources) ↑ based on a theory of meaning of frame semantics.

Frame semantics is the theory of linguistic developed by fill more it relates linguistic semantics to in sychometric knowledge. Here one can't understant the meaning of a single word without access to all the Essential knowledge that relates to that word.

For example: It is difficult to understand the meaning without knowing the situatio

i) identification of frames:

These process involves identification of frame invoked by predicates and creation of frame specific rolles called frame element.

Identification of predicate: which instanceted frame and labelled the sentences for those predicate.

For example: The concept of cooking involves a peson doing the cooking, the food that is want to cook, some thing to hold the food to cook, and source of container.

Frame: A frame is a script like conceptual structure that describes a particular type of situation, object or event along with the participants and propositions that are needed for the frame.

* Frame evoking words are called lexical units.

Frame Semantics: In frame semantics various types entities are involved like item, money, seller, buyer etc.

For example: The verb sell is used to exchange the item for money or its equivalent.

Relationship b/w entities:

* Buyer gives money to seller.
* Seller gives item to buyer
* Buyer believes value if item >= money amount
* Seller believes value of item <= monetary amount

Text Annotation: Text Annotation is a process in which we have an extract information about the words in a sentences. Ex: Bob donated the clock to John.

Uses:

1. Semantic roll assignment
2. Natural Language understanding (NLU)
3. machine translation
4. parts of speech tagging.
5. Information Extraction.

Prop bank: Prop bank is a corpus of text annotated with information about basic semantic propositions. It is created by Martha palmer. by adding predicate argument relation to the syntatic tree of a pen tree bank.

* The argument of each predicate are annotated with their semantic roles in relation to predicate the annotation requires the choice of senses id for each predicate an annotated with argument with semantic roles.
* propbank defines semantic roles on a verb by verb basis.

for example: [Devika]$_{ARG0}$ broken [the Window]$_{ARG1}$

In the given sentence devika is one argument and the window is another argument 1.

* [ARG 1 the window] Broke.
* Here the arguments of the verbs are labelled as numbered arguments.

Arg0, Arg 1, Arg2 ----- called core argument having label type (ARGN) or adjective argument having label type ARGM-X.

Tag	Description	Example
ARGM-LOC	Locative	The museum, in hyderabad
ARGM-TMP	Temporal	now, by nextweek.
ARGM-MNR	Manner	Heavily, clearly
ARGM-DIR	Direction	To market, to member
ARGM-MOD	Modal	Can, might, should, will

Semantic Role Labelling using SRL Algorithm.

Semantic role labelling also called as Shallow Semantic parsing in the process that assign label to words or phrases in a sentence that indicate their semantic role such as that of an agent, goal or result.

* It helps in the finding the meaning of the sentences.

For example: mary sold the book to john.

* In the given example mary is acting as a "agent" where as sold is a predicate the book is acting has a theme, the recipient is John.

* SLR is a unsupervised classification problem in which predicate & arguments of predicate are mapped to a node in a syntax tree.

* SLR consists of three tasks.

1. Argument Identification which involves of ~~valid~~ identification of valid semantic arguments of a predicate.
2. Argument classification which involves assging appropriate argument labels to identified constitances.
3. Argument identification and classification if it is a combination of above two tasks.

Uses:

Question answering, Information extraction, automatic text summarisation, text data mining, Speech recognination.

SRL ALGIORITHM

* Procedure: SRL (Sentences) returns best Semantic role Labelling.
* Input: Syntactic parse of the Sentence.
* Steps:

1. Generate a full syntactic parse of the Sentences.
2. Identify all the predicates.
3. For all predicates ∈ Sentences do
4. Extract a set of features of each node in the tree relative to the predication.
5. Classify each feature vector using the model created in training.
6. Select the class of highest scoring classifier.
7. Return best Semantic role labelling
8. End for

Pharse Structure Grammar:

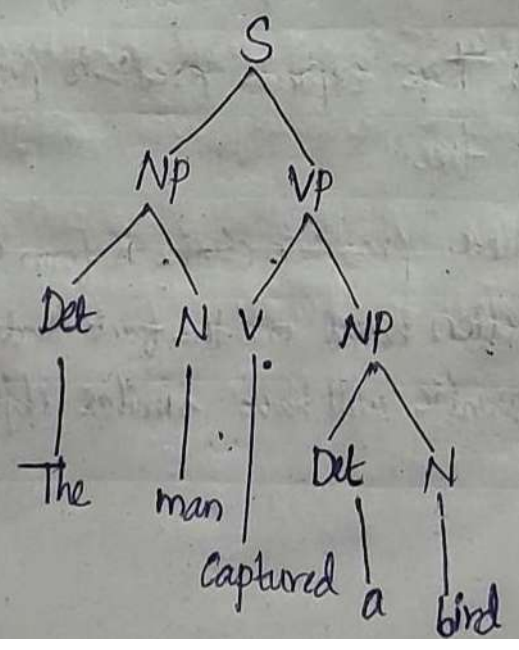
Phrase structure Grammar is a type of generative grammar in which constituent structures are represented by phrase structure rules or rewrite rules.

Rules:

S → NPVP

NP → Det N

VP → VNP



Generating a new structure parsing an structure

Phrase Structure Grammar for SRL problem:

In SRL a part from prop bank and framenet and Stastical Phrase can be used to generate a phrase structure tree.

* Daniel Gildea and Daniel Jurafsky ↓ used some features to address SRL problem.

The features are as follows:

1) path : It starts from argument 0 and goes upto predicate known as syntatic path from pause constituent to predicate.

   Ex: NP → S → VP

2) Predicate : this include identification of predicate lemma.

3) phrase type : this identifies types of phrase in a sentences licke

   EX: NP, VP, ADVP. etc..

4) position : It tells abaut the phrase position whether present in the beginning, or middle or at the end of predicate

5) voice : These features tells whether the predicate is in active or passive role.

6) Head word : These feature represented syntactic head of a Phrase and is calculated using head word table.

6) Sub categorization : These expands predicate parent node ina pause tree.

7) Verb clustering : Here predicate cluster is form using distance function based on the fact that verbs with similady Semantic will have similar objects.

Features suggested by Surdeanu et-al for SRL problem.

Content word: As head words were not so informative some rules set is used for finding the content word.

* If phrase type is PP then select the right most child.
  (phrase = "in Texas", content word = Texas)
* If phrase type is VP then select the head word.
  (phrase = had placed, content word = placed)
* If phrase type is ADVP then select the right most child.
  (phrase = more than, content word = "more")
* If the phrase is adjecent phrase then select the adj, verb, noun
  (phrase = 61 years old, content word = 61)
* For all other phrase type select the head word.
  (phrase = red house, content word = red)
* POS of the head word and the content word:
  These can increase the performance of decision tree system.
* name entity of the content word: These information can be represented as a set of binary values. (ARGM-TMP, LOC represents time and place entities)
* Boolean named entity flag: Entities like person, place, time, date etc. are considered as features.
* Phrase verb collection: The frquency of words immediately following preposition is computed and it is used as a feature.

Combinatory Categorical Grammar (CCG)

* In argument identification task, path feature is very important but it is difficult to trained those features.
* So with the help of CCG semantic role labelling can be improve. In these CCG three features are proposed.

1. Phrase type: which indicates the maximum projection the predicates and dependent word.
2. categorical path: These contains three values like dependent word categorie, direction of dependence and slot in the categorie filled by dependent word.
3. Tree path: It is a path from dependent word through predicate.

Chunk based approach for Semantic role Labelling:

* In case of speech data, chunk representation is faster and robust as compared to phrase wording.
* In chunk based system each base phrase is classified as beginning (B) inside (I) or outside (O) of a semantic role.
* This is known as IOB representation.
* In this approach the input text is first chunked or divided into base phrase by SVM classifier (support vector machine)
* In second step, SVM is trained for assignment of semantic labels to chunks.

Feature used by chunk based classifier.

Words	Words in the chunk
Postags	part of speech of the word in the chunk.
Named entities	The IOB tags of named entities
path	Defines a flat path b/w the token and the predicate.

Semantic chunker:

- Sentence → Chunk into Syntactic base phrase → Extract Feature → Generate parse

Nominalisation:

* It is the process when we take verbs or adjectives and turn them into nouns or nouns phrases.
* It prevent from repeating the same verb or other word again.

Ex: The government is introducing new tax rules that should help those on low incomes. (Verb)

Is introducing is a verb after nominalisation. The introduction of new tax rules should help those on low incomes

Verb nominalisation	
Verb	Noun
Achieve	Achievement
Adapt	Adaptation
perform	performance

Adjective Nominalisation	
Adjective	Noun
Angry	Anger
Available	Availability
Similar	Similarity

Predicate vs predicator:

* Predicate means the part of a sentences or clause containing a verb and stating something about the subject.
(Eg: John went home) where went home is a predicate and John is the subject.

$$\text{Predicate} = \text{Sentence} - \text{Subject}$$

* Predicator means a verb phrase considered as a constituent of clause structure along with subject object and adjunct".
* or verbal element of a sentences or clause.

$$\text{Predicator} = \text{verb}\ (\text{H.V} / \text{M.V} / \text{H.V} + \text{M.V})$$

(Helping verb or main verb)

Ex: My grand father was telling me stories.
H.V + M.V

* My book is here
Subject: My book
Predicate: is
predicator: here

Issues in SRL faced by languages other than English:

Following are some of the issues.

1. There are certain features which are present in some language.
Ex: No delimetters to separate the words in Chinese languages.
Due to these word semantic requires more complexes system as compared to English languages.

* Languages like chinese is morphological poor language. So predicate and arguments possess close connection as verbs nouns and adjectives have similarity.

* In chinese language syntactic parsing yeids less performance then english show shallow parse give more better results then full syntactic parse(or) passing.

3. In arbraic language pos categories are more in english or chinese mating it more challenging to handle all these by SRL system.

Meaning Representation System:

The biggest challenge faced by NLP system is the ambiguity which exists at every phrase It is a challenging task to take input in natural language, remove ambiguity use word knowledge and understand the context and make machines understand the language effortlessly like humans.

Resources Which are developed in this area include:

1. ATIS – Air Travel Information System

* It is a first system which transformed natural language (input information) ↑ for decision making by a application.

* It takes as an input user in speech Query about flight information. It converts into SQL Query & retrieve from flight database.

Here a semantic information encoded hierarchical frame representation.

Shows flight time
part of day:
origin: Hyd
Destination: Delhi
Date:
Day of journey: Tuesday

Natural representation

Communicator:

* It is an extension of ATIS (system).
* ATIS system was design for user initated dialogue i.e answers were provided by machine for user questions.
* communication system uses mixed initative dialogue that is it is an interactive system it is an communication happen b/w system & user. Based on real time information provided by the system user can itinary.
* many dialogues are collected by the system and stored in linguistic database about millions words & hundred (1600) dialogues are annotated.

Geo-Query:

* This system is designed for a domain of US geographic.
* It uses NLI (Natural Language Interface) to connect geographic database.
* geographic information like population, neighbouring states, major rivers and cities are stored in the relation DB. are stored in relation database.
* Sample Queries take the form as.

* Robocup : Clang.
* The Robocup coach Language was developed inable online coach these resource user robotic soccer as domain.
* C lang formal language is used which takes input from team coach and generate output in the form of behavior as if then rules.
  Ex: consider the advise by coach : if the ball is in our penalty area all our players, except player 4 should stay in our half.
* It is converted to rule.

* Ways of mapping Natural language to domain specifies meaning representation system (SQL) Query, prolog or domain specific (Query).
* Rule Based : Rule based semantic parsing system gives better results for ATIS and communicator system.
* Here speech recognation errors are handled by an interpreter.
* The system takes into account hierarchical frame structure and the value of these frames are adjust with few piece of information.

* Supervised : The short comings of rule base system like effects for creation of rules, time consuming, difficult to maintain, less scalable in case of complex problem.
* To overcome these difficulties end to end supervised statistical models with hand annotated data came into picture.

* The System has four components.

(a) Semantic parse

(b) Semantic frame

(c) Discourse

(d) Backend.