# **Develop an algorithm to extract syntactic cues and semantic labels from a narrative text such that they form an input to knowledge representation module.**

**“””Story chosen**: One day Buddha was walking through a village. A very angry and rude young man came up and began insulting him. “You have no right teaching others,” he shouted. “You are as stupid as everyone else. You are nothing but a fake.”

Buddha was not upset by these insults. Instead he asked the young man “Tell me, if you buy a gift for someone, and that person does not take it, to whom does the gift belong?”

The man was surprised to be asked such a strange question and answered, “It would belong to me, because I bought the gift.”

The Buddha smiled and said, “That is correct. And it is exactly the same with your anger.

If you become angry with me and I do not get insulted, then the anger falls back on you.**”””**

## Tokenization of the narrative text

*Tokenization* is the process of demarcating and possibly classifying sections of a string of input characters. The resulting tokens are then passed on to some other form of processing. The process can be considered a sub-task of [parsing](https://en.wikipedia.org/wiki/Parsing) input.

There are different types of tokenization on which sentence tokenization and word tokenization are some:

In Sentence tokenisation the whole story is fragmented into each sentences.

Sentence tokenization output:

['One day Buddha was walking through a village.', 'A very angry and rude young man came up and began insulting him.', '“You have no right teaching others,” he shouted.', '“You are as stupid as everyone else.', 'You are nothing but a fake.”Buddha was not upset by these insults.', 'Instead he asked the young man “Tell me, if you buy a gift for someone, and that person does not take it, to whom does the gift belong?”The man was surprised to be asked such a strange question and answered, “It would belong to me, because I bought the gift.”The Buddha smiled and said, “That is correct.', 'And it is exactly the same with your anger.If you become angry with me and I do not get insulted, then the anger falls back on you.']

In Word tokenisation the whole story is fragmented into each words.

Word tokenization output:

['One', 'day', 'Buddha', 'was', 'walking', 'through', 'a', 'village', '.', 'A', 'very', 'angry', 'and', 'rude', 'young', 'man', 'came', 'up', 'and', 'began', 'insulting', 'him', '.', '“', 'You', 'have', 'no', 'right', 'teaching', 'others', ',', '”', 'he', 'shouted', '.', '“', 'You', 'are', 'as', 'stupid', 'as', 'everyone', 'else', '.', 'You', 'are', 'nothing', 'but', 'a', 'fake.', '”', 'Buddha', 'was', 'not', 'upset', 'by', 'these', 'insults', '.', 'Instead', 'he', 'asked', 'the', 'young', 'man', '“', 'Tell', 'me', ',', 'if', 'you', 'buy', 'a', 'gift', 'for', 'someone', ',', 'and', 'that', 'person', 'does', 'not', 'take', 'it', ',', 'to', 'whom', 'does', 'the', 'gift', 'belong', '?', '”', 'The', 'man', 'was', 'surprised', 'to', 'be', 'asked', 'such', 'a', 'strange', 'question', 'and', 'answered', ',', '“', 'It', 'would', 'belong', 'to', 'me', ',', 'because', 'I', 'bought', 'the', 'gift.', '”', 'The', 'Buddha', 'smiled', 'and', 'said', ',', '“', 'That', 'is', 'correct', '.', 'And', 'it', 'is', 'exactly', 'the', 'same', 'with', 'your', 'anger’,’.’,If', 'you', 'become', 'angry', 'with', 'me', 'and', 'I', 'do', 'not', 'get', 'insulted', ',', 'then', 'the', 'anger', 'falls', 'back', 'on', 'you', '.']

* 1. Develop an algorithm to extract the syntactic cues from the text.

A paragraph is given and there should some information in the syntactic structure of the given data. As humans we find that task easy until when we think of that being thought to the machine.

In this section with the help of the syntactic rules/information in the given text is first analysed and the grammar is set to extract only the required words that will get the overview of the story.

**Algorithm :**

* Read text.
* Identify the Special Characters and flag them(Not included for POS tagging)
* Creation of the dictionary the only picks the required words
  + Words that starts with the Noun=>Adjective=>Noun
  + Words ending (Leaf Ending) which has a conjunction as parent.
* Perform tokenisation eliminating the special characters.
* Run through parts of speech tagger.
* With the help of the dictionary only select the words that fit into and print.

**Output:**

Table 2.1‑1:Output from the syntactic extractor

|  |
| --- |
| **Sentence 1** |
| day |
| buddha |
| village |
| **Sentence 2** |
| rude |
| young |
| man |
| **Sentence 3** |
| right |
| teach |
| other |
| **Sentence 4** |
| everyone |
| **Sentence 5** |
| nothing |
| fake |
| buddha |
| insult |
| **Sentence 6** |
| young |
| man |
| tell |
| gift |
| someone |
| person |
| gift |
| belong |
| man |
| strange |
| question |
| **Sentence 7** |
| gift |
| buddha |
| anger |
| anger |

Tabel 2.1-1 Represents the extraction of the key words that determines the story. In the above table it is observed that the word obtained are of the name of the people, things and places are recognised and extracted.

* 1. Formation of a semantic label table

Semantics is the meaning of words in sentences. it refers to meaning that does not depend on the context where it appears. Semantic table consists of [Context1, Intent1, Dependencies, Dep].

Context has the words in the sentence ,Intent represents to the Action that is performed in the sentence, Dependencies represents the dependence of other words in the same sent to the word in the context, Dep is the semantic role label.

Semantic label table of the first Sentence.

0 One day [] nummod

1 day walking [One] npadvmod

2 Buddha walking [] nsubj

3 was walking [] aux

4 walking walking [day, Buddha, was, through, .] ROOT

5 through walking [village] prep

6 a village [] det

7 village through [a] pobj

8 . walking [] punct

## Development of an algorithm to extract the semantic labels

Semantic labelling tells the meaning of the text given as the input. This is done with the following steps. In this section the sematic label is considered as the Intent Recognition. Hence the Intent tells the whole action that is happening in the sentence.

**Algorithm:**

* Read text.
* Tokenize the paragraph into sentences.
* Identify the ‘and’ and ‘also’ words in the input and combine if wrongly tokenised in tokenisation.
* Links between the words are identified.
* The word with maximum links is the word that contains the action(Meaning) of the sentences.
* The root word is extracted from the whole sentence.

This Root word gives the picture of what happened in that line.

Tabel 2.4-1 consists of the words that represent the actions that is performed in each of the sentences.

**Output:**

Table 2.4‑1:Output of the semantic Extractor

|  |  |
| --- | --- |
| Sentence 1 | Walking |
| Sentence 2 | Angry, Came, Began |
| Sentence 3 | Shouted |
| Sentence 4 | Are Stupid |
| Sentence 5 | Are Fake |
| Sentence 6 | Take, gift, Buy, Belong |
| Sentence 7 | Falls, Anger |

## Development of an algorithm for coreference resolution for the extracted syntactic and semantic information

Coreference resolution is the task of determining linguistic expressions that refer to the same real-world entity in natural language.

The Coreference is done with the help of nueralCoref which gives the graphical coreference of the story.

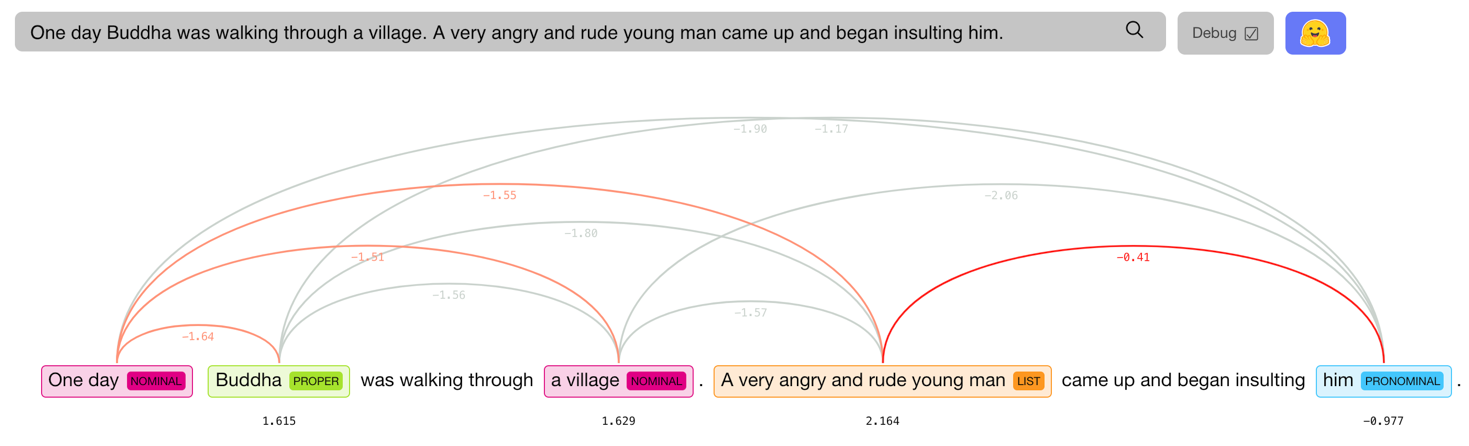


Figure 2.5‑1:Output of the coreference Coefficient of the first two sentences.

Figure 2.5-1 shows the coreference coefficient of the first two sentences and from which the relationship of the words is evaluated as coefficient values.

With the output of Syntactic and semantic information obtained fed into the Coreference Coefficient calculator it gives the relationship between the important words obtained from the syntactic extractor and the words from the semantic extractor.

## Feature identification by integrating syntactic, semantic, and coreference information

Each every step that is achieved until now has given some important information for further processing to be precise with the information contributed by the individuals are as follows.

* **Syntactic Extractor: Base words that form the story.**
* **Semantic Extractor: Main Theme of the sentence.**
* **Coreference resolution: Relationship between linguistic expression and real world word.**

By combining the output of the three previous steps sentence wise it gives an model that is good for the analysis purpose.

Syntactic Extractor: gives out all the person ,things and sometimes feelings from the story.

Semantic Extractor: gives the happening I.e. the root/cause.

Coreference coefficient: helps in framing he, him, that, etc.

The Syntactic and semantic words considered as the nodes of the model. The Conditional probability distribution is assigned to the Semantic extractor nodes. These nodes are constructed sentence wise.

# **PART-C CHAPTER 3**

\_\_\_\_­­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# **Develop an algorithm for action description of a narrative text using Bayesian network**

## Development of a knowledge representation algorithm using the features identified in Part-B.

* The actual story the main context of the story lies in the conversation of Buddha and the young Man.
* The part before the conversation tells the thinking of the young man about Buddha.
* The Conversation refers to the Intent call gift but ultimately it is referred to the anger and the hurt.
* The complete reference to gift has to be taken to the Anger and Hurt.

## Identification of threshold patterns for joint probability distribution

As the story is seen in perspective of two parts I.e.

* Before conversation
* While conversation

Before Conversation the nature of Buddha is explained

While Conversation about Purchase and acceptance of Gift is referred

This demands us to model the problem with two different graphical model for before and while conversation.

## Development of a Bayesian network model

The Bayesian Network model is designed as per the identification of the threshold patterns Hence there are two Bayesian network graphs.

* Bayesian Graph for before conversation That represents the Nature of Buddha.

In this part the Ability of Buddha is inferenced weather he is a valid teacher or not with the quality that was mentioned by the young man.

* Bayesian Network for while conversation the dictates about gift that ultimately means Anger and Hurt.

In this part the acceptance of Anger and Hurt is expressed with the intent Gift

## Development of a reasoning scheme for inferring action description

Reasoning scheme for inferring the action is done using the python library for probabilistic graphical model **pgmpy**.