

**TITLE:** Computational Linguistics I  
**COURSE CODE:**  
**CREDITS:** 3-1-0-4  
**Type When:** Summer 2021 (Spring T3 2021)  
**FACULTY:** Radhika Mamidi  
**PRE-REQUISITE:** None

**OBJECTIVE:** To provide students with the basic understanding of language analysis based on linguistic theories and show them how these can be applied for NLP. The course will familiarise the students with the linguistic challenges in processing natural languages, analyzing the structures and dealing with ambiguities in language.

## **COURSE TOPICS:**

**1. INTRODUCTION TO THE COURSE: What is CL and where does it apply? Issues and challenges**

**2. Language processing pipeline for text processing: Structural Analysis at various levels –** word (POS, morphology), phrase (chunk), sentence (syntactic parsing). Word meaning: Lexical Semantics, Dealing with Ambiguities (WSD/WTB)

**2.1. MORPH ANALYSIS:** Morph analysers and word generators

Recap of basic units in word formation: morphemes, allomorphs. Word formation: Affixation, suffixation, prefixation, infixation; Non-concatenative, Compounding, Morphotactics; Constraints on affixes; Morphophonology. Types of word formation processes (function based): inflectional, derivational.

Developing morph analysers and generators: finite state automata, paradigm tables, add-delete rules.

**2.2. WORD MEANING:** Lexical semantics

Hypernymy, hyponymy, synonymy, antonymy, lexicon and lexicography; machine readable dictionaries, WordNet, ConceptNet, VerbNet etc.

**2.3. SHALLOW PARSING AND SENTENCE PARSING:** Words and their arrangements in a sentence.

### **2.3.1 POS Tagging**

Word classes, Parts of Speech, POS tagging, Rule based parts of speech taggers, Statistical parts of speech taggers, Annotating POS tagged data, Issues in tagging, Defining tagset for your languages.

### **2.3.2 Shallow parsing (arrangement of words in a sentence)**

#### **2.3.2.1 Local Word Grouping (LWG)**

Grouping functional words such as prepositions/postpositions and auxiliaries with the content words (nouns, verbs).

#### **2.3.2.2 Chunking:** Forming minimal phrases

**2.4. MULTIWORD EXPRESSIONS (MWEs):** Named entities (NEs), Idioms, compounds. Types of named entities; compositionality in MWEs.

## **2.5. SYNTACTIC PARSING: Analysing the structure of a sentence, grammatical approaches**

### **2.5.1 Constituency Analysis:** Constituents/ phrases.

Deriving sentences using phrase structure rules (CFG); Constraints on rules; Subcategorization; verb argument structure. Representing phrase structures: X-bar schema, Complements and Adjuncts. Syntactic operations: Substitution, adjunction and movement. Syntactic phenomena: Passive, Raising, Control.

### **2.5.2 Dependency Analysis:** Dependency structures: Head – modifier relations.

Paninian grammar – a dependency framework – relations in Paninian grammar: karaka, tadarthya, hetu etc; Vibhakti - relation marker; karaka vibhakti mapping, karaka chart

### **2.5.3 Parsing approaches:** English parsers, Hindi/IL parsing using Paninian framework.

## **3. Speech Processing:**

Introduction to speech processing: Speech production, Speech perception,  
Speech analysis  
Speech Recognition  
Speech Synthesis

### **PREFERRED TEXT BOOKS:**

1. Jurafsky & Martin, 2000; Speech and Language Processing, Pearson Education
2. Bharati et al., 1995; Natural Language Processing: A Paninian Perspective
3. Fundamentals of Speech Recognition by Lawrence Rabiner, Biing-Hwang Juang
4. The Oxford Handbook of Computational Linguistics. 2003. Ruslan Mitkov (ed)

### **REFERENCE BOOKS :**

Fromkin, V, Robert Rodman, Nina Hyams (2002) *An Introduction to Language*, Thomson Wadsworth

Fromkin, V (editor) (2003) *Linguistics: An introduction to Linguistic Theory*

Aronoff, Mark and Janie Rees-Miller (eds) (2003), *The Handbook of Linguistics*, Blackwell Publishers

**PROJECT:** The course will have a project content where students will study and solve a problem using real language data.

**GRADING:** Graded exercises 20%, Assignments 40%, Project 20%, Seminar 20%

**OUTCOME:** At the end of the course the students will be able to understand and analyse actual language data and develop computational resources for various levels of language structures