



**STANLEY**  
**COLLEGE OF ENGINEERING & TECHNOLOGY FOR WOMEN**  
( Approved by AICTE , New Delhi | Affiliated to Osmania University ,Hyderabad)  
Address : Chapel Road, Abids ,Hyderabad

**CS 565**

## **NATURAL LANGUAGE PROCESSING**

### **UNIT-I**

Introduction  
Elementary Probability Theory  
Essential Information Theory

### **UNIT-II**

Linguistic Essentials Corpus-Based Work

### **UNIT-III**

Collocations.

**Statistical Inference:** Bins: Forming Equivalence Classes, Reliability vs. Discrimination, n-gram models, Building n-gram models, An Information Theoretic Approach.

**Word Sense Disambiguation:** Methodological Preliminaries, Supervised and unsupervised learning, Pseudo words. Upper and lower bounds on performance, Supervised Disambiguation, Bayesian classification.

### **UNIT-IV**

Evaluation Measures

**Markov Models:** Hidden Markov Models, Use, General form of an HMM Part-of-Speech Tagging

### **UNIT-V**

**Probabilistic Context Free Grammars:** Introduction

Clustering

**Information Retrieval:** Background, The Vector Space Model

### **Suggested Reading:**

1. Christopher D. Manning, Hinrich Schütze, "Foundations of Statistical Natural Language Processing", The MIT Press, 1999.
2. James Allan, "Natural Language Understanding", Pearson Education, 1994.
3. Tanveer Siddiqui, US Tiwary, "Natural Language Processing and Information Retrieval", Oxford University Press, 2008.