



SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 97

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

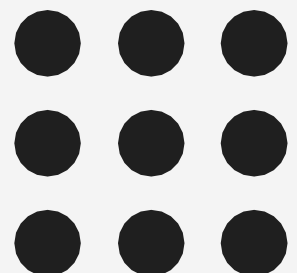
Department of Artificial Intelligence and Data Science

**Course Name – 19AD601 – Natural Language
Processing**

III Year / VI Semester

Unit 2 – WORD LEVEL ANALYSIS

Topic 7- Issues in Pos Tagging





Issues in Pos Tagging



Ambiguity

The main problem with POS tagging is ambiguity.

Ambiguity issue arises when a word has multiple meanings based on the text and different POS tags can be assigned to them.

In English, many common words have multiple meanings and therefore multiple POS. The job of a POS tagger is to resolve this ambiguity accurately based on the context of use.

Example: Take a word “can” which has multiple tags

can/verb can/auxiliary can/noun

Average of ~2 parts of speech for each word



Issues in Pos Tagging

Types:		WSJ		Brown	
Unambiguous	(1 tag)	44,432	(86%)	45,799	(85%)
Ambiguous	(2+ tags)	7,025	(14%)	8,050	(15%)
Tokens:		WSJ		Brown	
Unambiguous	(1 tag)	577,421	(45%)	384,349	(33%)
Ambiguous	(2+ tags)	711,780	(55%)	786,646	(67%)

The number of tags used by different systems varies a lot. Some systems use < 20 tags, while others use > 400.



THANK YOU