

Using Language Models for POS Tagging

Training data (Annotated text)

<i>This</i>	<i>sentence</i>	<i>serves</i>	<i>as</i>	<i>an</i>	<i>example</i>	<i>of</i>
Det	N	V1	P	Det	N	P
<i>annotated text...</i>						
V2	N					

“This is a new sentence”



This is a new sentence
 Det Aux Det Adj N

Consider all possibilities, and pick the one with the highest probability

<i>This</i>	<i>is</i>	<i>a</i>	<i>new</i>	<i>sentence</i>
Det	Det	Det	Det	Det
...	...			
Det	Aux	Det	Adj	N
...	...			
V2	V2	V2	V2	V2

$$p(w_1, \dots, w_k, t_1, \dots, t_k) \\
 = \begin{cases} p(t_1 | w_1) \dots p(t_k | w_k) p(w_1) \dots p(w_k) \\ \prod_{i=1}^k p(w_i | t_i) p(t_i | t_{i-1}) \end{cases}$$

Method 1: Independent assignment
 Most common tag

Method 2: Partial dependency

$w_1 = \text{“this”}, w_2 = \text{“is”}, \dots, t_1 = \text{Det}, t_2 = \text{Det}, \dots$