

# Note of Natural Language Processing

## Preliminaries and Modern Models

Junjie Liu<sup>1\*</sup>  
liuj13@tcd.ie

<sup>1</sup> Department of Political Science, Trinity College Dublin

Date of Final Compile: November 29, 2024

### Contents

<b>1</b>	<b>Preliminaries</b>	<b>3</b>
<b>2</b>	<b>Foundation Models</b>	<b>3</b>
<b>3</b>	<b>Large Language Models</b>	<b>3</b>

---

\*Considered of my limited ability, mistakes are inevitable. Correction and suggestions are welcomed. This work is under the CC-BY-SA 4.0 International license

## Abstract

In this note, we are not trying to do the systematic review of the entire NLP development, however, it would be the collection of some basic knowledge and modern models in the field of NLP. The note is divided into three parts: preliminaries, foundation models, and large language models. The first part will introduce some basic concepts and methods in NLP, such as tokenization, POS tagging, and parsing, also the related mathematics basis of NLP. The second part will introduce some foundation models in NLP, such as Hidden Markov Model, Maximum Entropy Model, and Conditional Random Field. The third part will introduce some large language models, such as T5, LLaMA, and, GPT. The note is not intended to be comprehensive, but to provide a brief overview of the field of NLP to the scholar that trying to use text-as-data in their research.

- 1 Preliminaries**
- 2 Foundation Models**
- 3 Large Language Models**

## References