

**Differences Between The Uses of The Character 道 in Confucian And Daoist Canon: A  
Reflective Essay**

Bowen Li

Department of Language Science and Technology, The Hong Kong Polytechnic University

CHC5904: Digital Studies of Chinese Culture

Dr. Jing CHEN

Oct 27, 2025

# Differences Between The Uses of The Character 道 in Confucian And Daoist Canon: A

## Reflective Essay

### Introduction

“Tao Te Ching” and “The Analects of Confucius” are two traditional Chinese philosophical canons from Daoism and Confucianism. Moreover, the character “道” is one of the frequently used terms in Chinese philosophical classics (道, 法, 礼, 德, 仁, 义, 无, 有, 名, 天). And it is widely used in these two canons. Therefore, this article studies the differences in the usage of the character “道” in these two different canons and, by extension, the two different schools of thought, from the perspectives of word frequency and its high-frequency collocations (nearby frequently co-occurring words).

### Workflow

Since the workflow of each canon is the same, the figure is randomly chosen from the two canons for reference.

### Data Scraping

The data is scraped from Ctext.org (<https://ctext.org/analects/zh> & <https://ctext.org/dao-de-jing/zh>) by manual data scraping. And the raw data in .txt format is demonstrated as Figure 1 and 2. There are 81 and 504 paragraphs in the two canons, respectively.

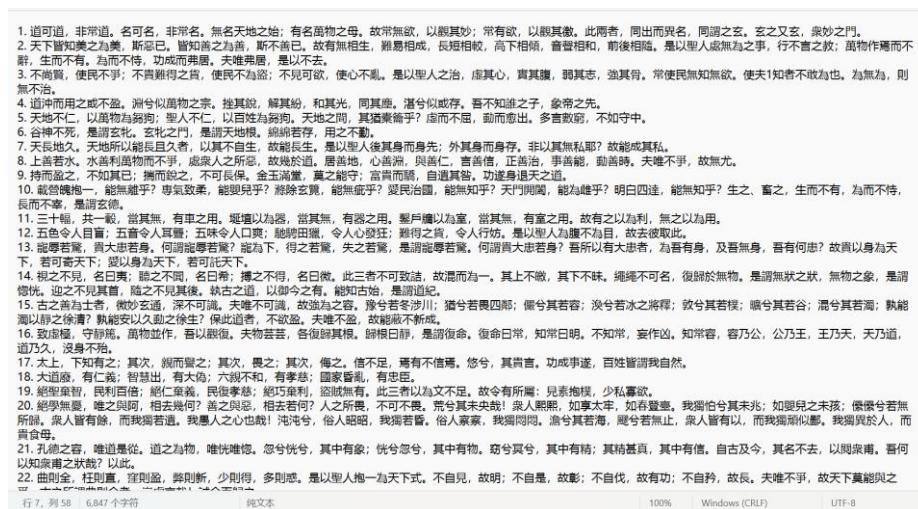


Figure 1

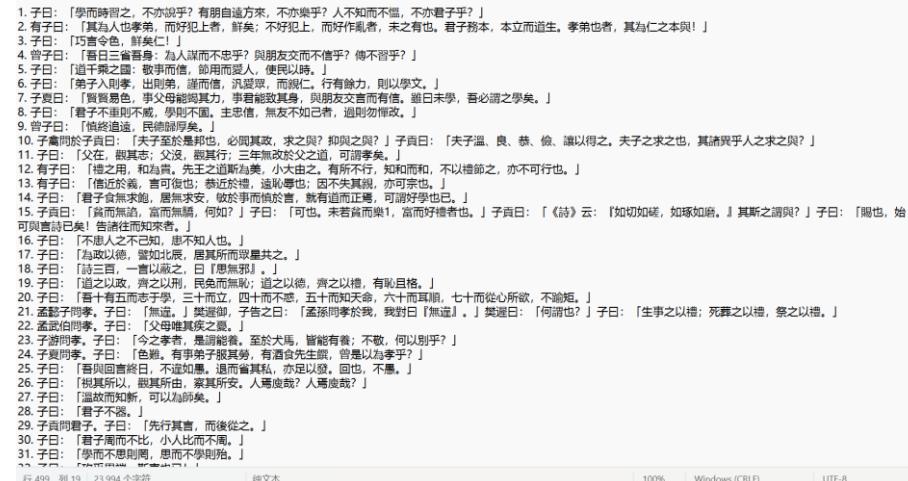


Figure 2

## Data Cleansing

By loading the .txt file into *AntConc*, and clicking “Word” button to calculate the frequency of the character “道”, it is found that the frequency of the single character cannot be illustrated, as Figure 3 shows.

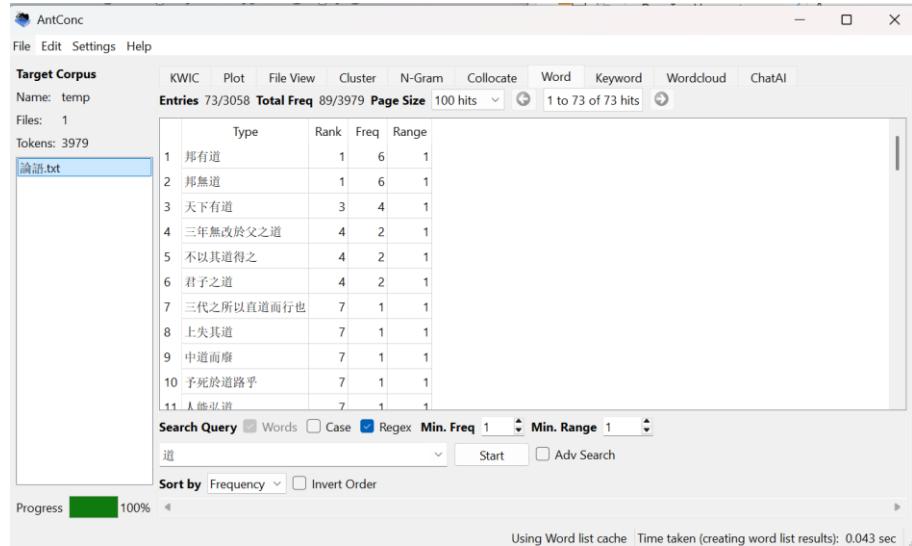


Figure 3

It is because there is no blank between each character, so that only Cases or Regex function can be used rather than Words function. Therefore, a Python code (Figure 4) package is used to add blank in each character so that it can operate smoothly. And the frequency of the character is calculated successfully, as Figure 5 shows.

```

1 < with open('論語.txt', 'r', encoding='utf-8') as f:
2 |   content = f.read()
3 |
4   result = ''
5 < for char in content:
6 |   if char not in [' ', '\n', '\r', '\t']:
7 |     result += char + ' '
8 <   else:
9 |     result += char
10
11 < with open('論語_分词版.txt', 'w', encoding='utf-8') as f:
12 |   f.write(result)
13
14 print("处理完成! ")

```

Figure 4

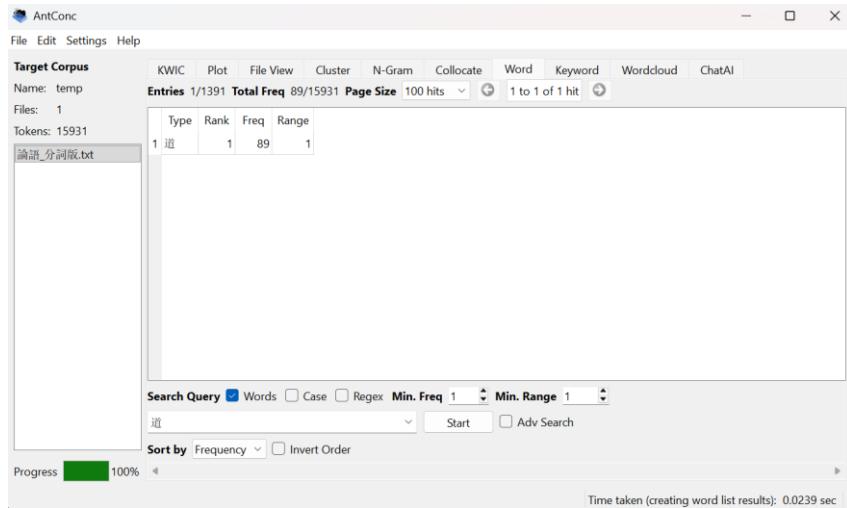


Figure 5

## Data Visualization

After clicking the Cluster function and ticking the On Left/Right button, the characters that coexists with “道” are demonstrated, as Figure 6 shows. And by a Python code package, the data is transferred to .csv format for another Python code package to generate word cloud, as Figure 7 shows.

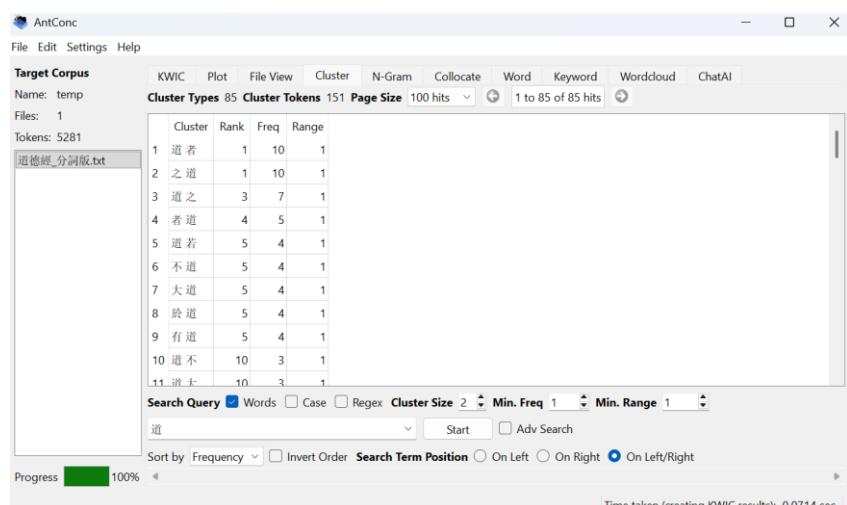


Figure 6



*Figure 7*

More detailed coding package and other documents, please see GitHub link in the PPT slide in the later pages.

## Reflection of Using DH Tools

The process of examining the character “道” in the Tao Te Ching and The Analects using digital tools like AntConc and Python has been both enlightening and transformative for my understanding of textual analysis in the digital humanities. Using AntConc for corpus analysis proved particularly valuable for identifying frequency patterns and clusters. The KWIC function allowed me to observe how “道” operates in different syntactic environments across the two texts for deeper understanding. However, during practical execution, it is found that two different dataset cannot be loaded and manipulated in AntConc simultaneously, which might be harder for users to make comparison jobs between two datasets.

The Python workflow, from text preprocessing to visualization, presented both challenges and rewards. The initial step of adding spaces between characters—crucial for accurate word segmentation in classical Chinese—required careful handling of punctuation and special characters. Generating the word cloud, while visually compelling, also highlighted the limitations of frequency-based visualizations in capturing the nuanced philosophical meanings of “道” since the Chinese font seemed not available in AntConc so only some squares are illustrated on the Wordcloud function it originally has. The most valuable aspect of the Python implementation was creating structured CSV files that enabled systematic comparison of frequency data between the two texts that the visualization part offers.

# Differences Between The Uses of The Character 道 in Confucian And Daoist Canon

Hands-on Assignment #1



GitHub Repository:

<https://github.com/samklein112/Differences-Between-The-Uses-of-The-Character-in-Confucian-And-Daoist-Canon>

Li Bowen  
25052859G

Department of Language Science and Technology





# Content

- **Introduction**
    - ✓ Research Questions Proposal
  - **Methodology**
    - ✓ Methodology and Tools
  - **Findings and Results**
    - ✓ Findings and Results towards RQs
  - **Reflection**
    - ✓ Answer to RQs
- 

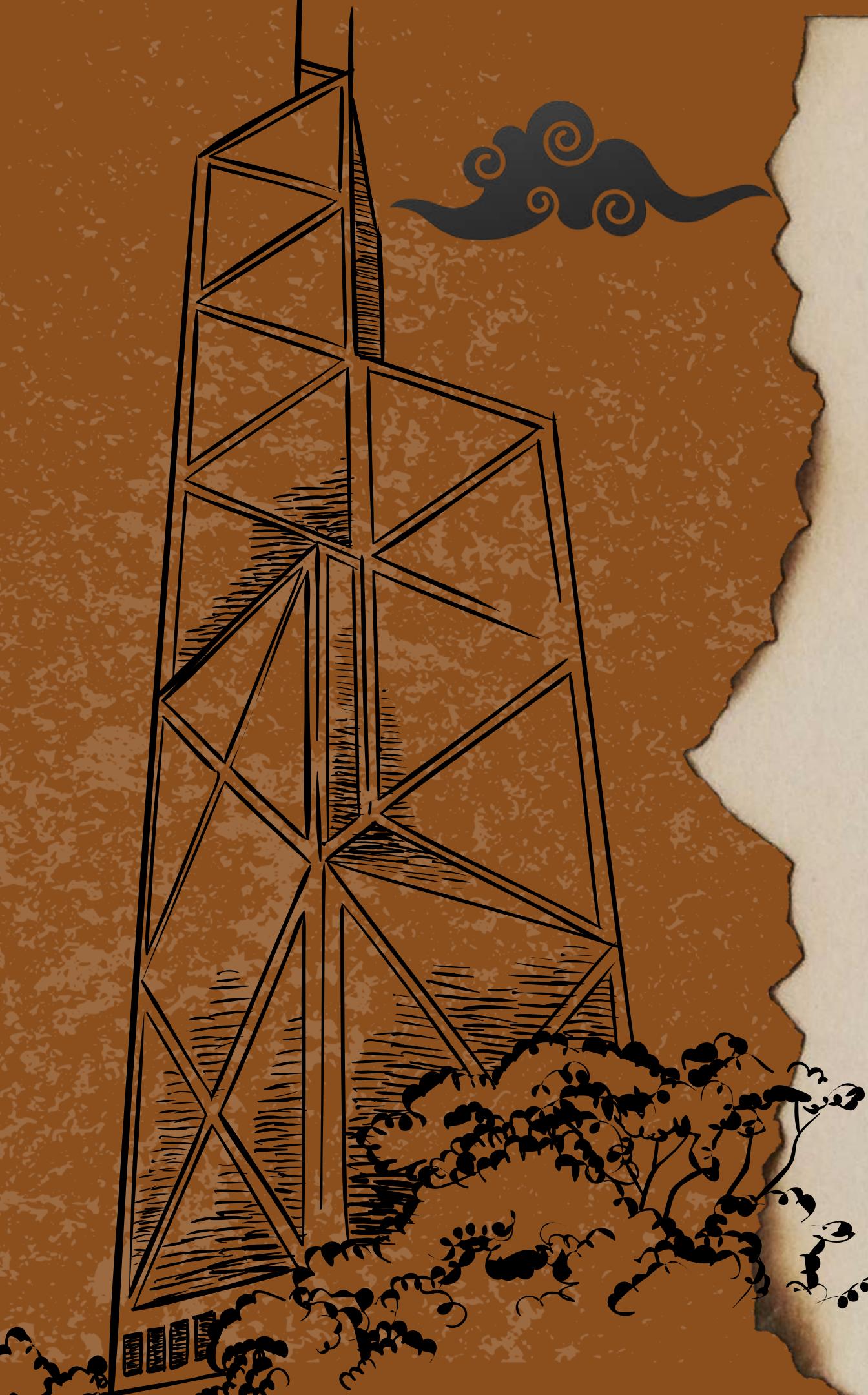
# Introduction

- This research studies the differences in the usage of the character “道” in these two different canons and, by extension, the two different schools of thought, from the perspectives of word frequency and its high-frequency collocations (nearby frequently co-occurring words) **with the following research questions:**

**RQ1: How does the relative frequency of the character “道” differ between the two canons?**

**RQ2: How do the most frequent clusters of “道” reveal its distinct semantic and philosophical roles in each canon?**





# Methodology

- The corpus linguistics approach is used, which involves the computer-assisted analysis of language patterns within a body of text (a "corpus").

The specific methodologies used were:

**Word Frequency Analysis:** To determine the absolute and relative prominence of the keyword “道”.

**Cluster Analysis:** To identify the most frequent word pairs (cluster) containing “道”. This reveals its common linguistic contexts and conceptual partnerships.

**Key Word in Context (KWIC):** To examine individual concordance lines and understand how “道” is used in specific sentences, providing qualitative support for the quantitative data.



# Tool

- The corpus linguistics approach is used, which involves the computer-assisted analysis of language patterns within a body of text (a "corpus").

The specific tools used were:

**AntConc:** The primary tool for corpus analysis. It was used to generate all frequency lists, cluster tables, and KWIC analysis.

**Python:** Used for initial data processing and file preparation to ensure compatibility with AntConc.

# Findings & Results

## Finding #1: Differential Prominence of “道”

- While *The Analects of Confucius* uses the character “道” more often in raw count (89 times vs. 76 times), *Tao Te Ching* gives it significantly more relative prominence. The *Tao Te Ching* is a much shorter text (5,281 tokens) compared to *The Analects of Confucius* (15,931 tokens).

In *Tao Te Ching*, “道” appears 76 times per 5,281 tokens ( $\approx 1.44\%$  of the text).

In *The Analects of Confucius*, “道” appears 89 times per 15,931 tokens ( $\approx 0.56\%$  of the text).

This simple frequency normalization shows that “道” is a more densely concentrated and central term in the *Daodejing*, appearing over **2.5 times more frequently** relative to text length.



# Findings & Results

## Finding #2: Distinct Semantic Profiles Revealed by Cluster Analysis

- The **cluster analysis** (identifying frequent word pairs) reveals starkly different conceptual networks surrounding “道” in each text, which is powerfully illustrated by examining the original texts.



# Top Frequent Clusters in *The Analects of Confucius*



# Top Frequent Clusters in *Tao Te Ching*



# Findings & Results

## Tao Te Ching's “道”: A Cosmological and Generative Force

The clusters portray “道” as the fundamental, generative principle of the universe. It is often the **subject** of descriptive or active sentences.

大道: “大道汎兮，其可左右。” (Ch. 34) *Indicates its supreme, all-encompassing nature.*

道生: “道生一，一生二，二生三，三生萬物。” (Ch. 42) *This is the ultimate generative act, establishing Dao as the origin of all existence.*

天道: “天道無親，常與善人。” (Ch. 79) *Describes Dao as an impartial, natural force governing the cosmos.*

## The Analects’s “道”: A Socio-Ethical and Political Path

The clusters show “道” as a path or standard that people and states can either follow or deviate from. It is often the **object** of action or discussion.

有道 / 無道: “天下有道則見，無道則隱。” (8.13) “道” is a condition of the world, something that can be present or absent, determining one's action.

道之: “道之以政，齊之以刑...” (2.3) It is a method of governance, an action performed upon the people.

道也: “夫子之道也，忠恕而已矣。” (4.15) “道” is defined in human, ethical terms—it is a teachable doctrine consisting of specific virtues.



# Reflection

In answer to the research questions:

1. **The *Tao Te Ching* gives far more relative prominence to “道” than *The Analects*.** The frequency analysis confirms that “道” is the absolute, central subject of the *Tao Te Ching*, appearing as its conceptual core. In contrast, while equally fundamental, “道” in *The Analects* is one key virtue among others (e.g., 仁 *benevolence*, 礼 *ritual*) within a broader socio-ethical framework.
2. **The cluster analysis reveals fundamentally different conceptions of the term.** The semantic profiles built from the top clusters provide a clear, data-driven answer:
  1. In the *Tao Te Ching*, “道” is a **cosmological and generative force**. Clusters like 大道, 道生, and 天道 portray it as the impersonal, creative source of the universe—a metaphysical reality to be described and understood.
  2. In *The Analects*, “道” is a **socio-ethical and political path**. Clusters like 有道/無道 and 道之 frame it as a tangible standard for human conduct and governance—a moral ideal that can be followed, possessed, lost, or applied.

# *Thank You*

Li Bowen

25052859G

Department of Language Science and Technology

