

New Perspectives on the Classical Chinese Philosophy: A Computational Approach

Co-PIs: James Miller & Charles Chang

Student Leader: Junyi Tao

Student Team:

Distant Reading: Youzhang He | Xiaoyin Chen

Close Reading: Miechen Liu | Jiaming Li



Research Objectives

- Exploring ways to capture the distinctive features of the philosophical schools of the pre-Qin and Han dynasties
- Long-term goal: track the change over time



A decorative graphic of a leafy branch, rendered in a light blue line-art style, extending from the left edge of the slide across the top and middle sections.

How to capture the features of schools

CONTENT

- meanings of polysemous words varying across schools
- their distributions and themes they represent

WRITING STYLE

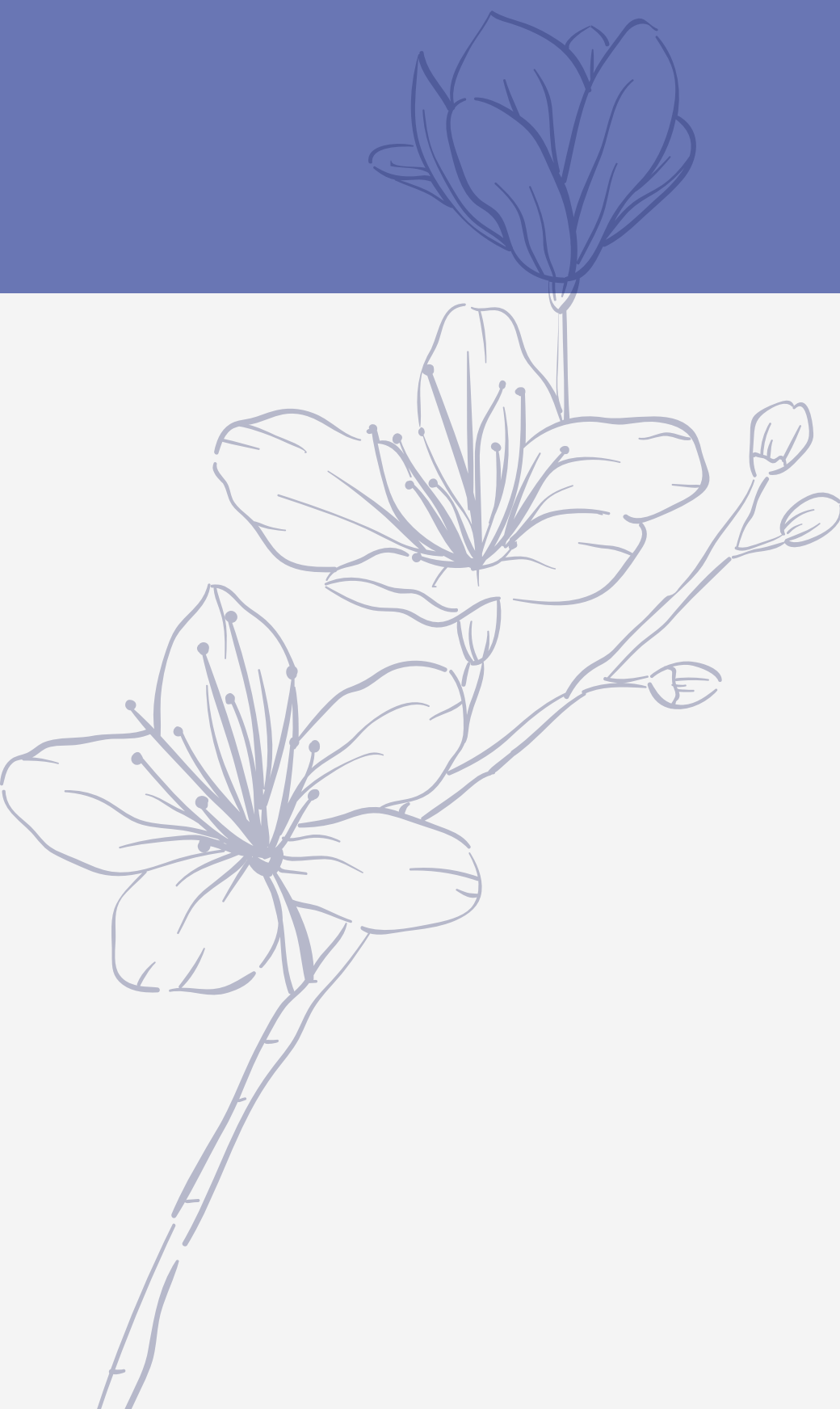
- frequency of entities
- sentence structure

Motivation

To discover perspectives that cannot be illustrated by the traditional close reading paradigm

1. Using the agnostic nature of algorithms
2. Large-scale analysis

Exploratory!



Methods

1. Close reading: summarize patterns + interpretations
2. Distant reading
 - a. Unsupervised: provide suggestive meanings/patterns
 - b. Supervised: test the importance of selected features
3. Close <--validate--> Distant

Corpus Description

Chinese Text Project

Welcome

Welcome to the Chinese Text Project homepage. The Chinese Text Project is an online open-access digital library that makes pre-modern digital medium to explore new ways of interacting with these texts that are not possible in print. With over thirty thousand titles and existence.

You may wish to [read more about the project](#), view the [pre-Qin and Han](#), [post-Han](#) or [Wiki](#) tables of contents, or [consult the instructions](#) main textual sections of the site.

網站有中、英文版本，也有繁、簡體版，可通過每頁左上角的連結隨時調整。

This site is provided as an open-access resource. **If you use this site in your teaching or research, please ask your university**

<https://ctext.org/>

The Chinese Text Project is an online open-access digital library that makes pre-modern Chinese texts available to readers and researchers all around the world. With over thirty thousand titles and more than five billion characters, the Chinese Text Project is also the largest database of pre-modern Chinese texts in existence.

Pre-Qin and Han	Confucianism	先秦两汉	儒家
	Mohism		墨家
	Daoism		道家
	Legalism		法家
	School of Names		名家
	School of the Military		兵家
	Mathematics		算书
	Miscellaneous Schools		杂家
	Histories		史书
	Ancient Classics		经典文献
Etymology		字书	
Chinese Medicine		医学	
Excavated texts		出土文献	
Post-Han	Wei, Jin, and North-South	汉代之后	魏晋南北朝
	Sui-Tang		隋唐
	Song-Ming		宋明
	Qing		清代
	Republican era		民国

March 24, 2023



Content Analysis — meanings



You need a theory of the meaning of concepts
And a theory of the words and phrases that express them in the corpora

What is the meaning of Tescüino?

Some contexts:

1. A bottle of Tescüino is on the table
2. Everybody likes Tescüino
3. Don't have Tescüino before your drive
4. We make Tescüino out of corn



JOHN RUPERT FIRTH

Distributional Hypothesis: You shall know a word by the company it keeps (Firth, 1957)

=> need to represent the contexts, e.g. collocations or embeddings...

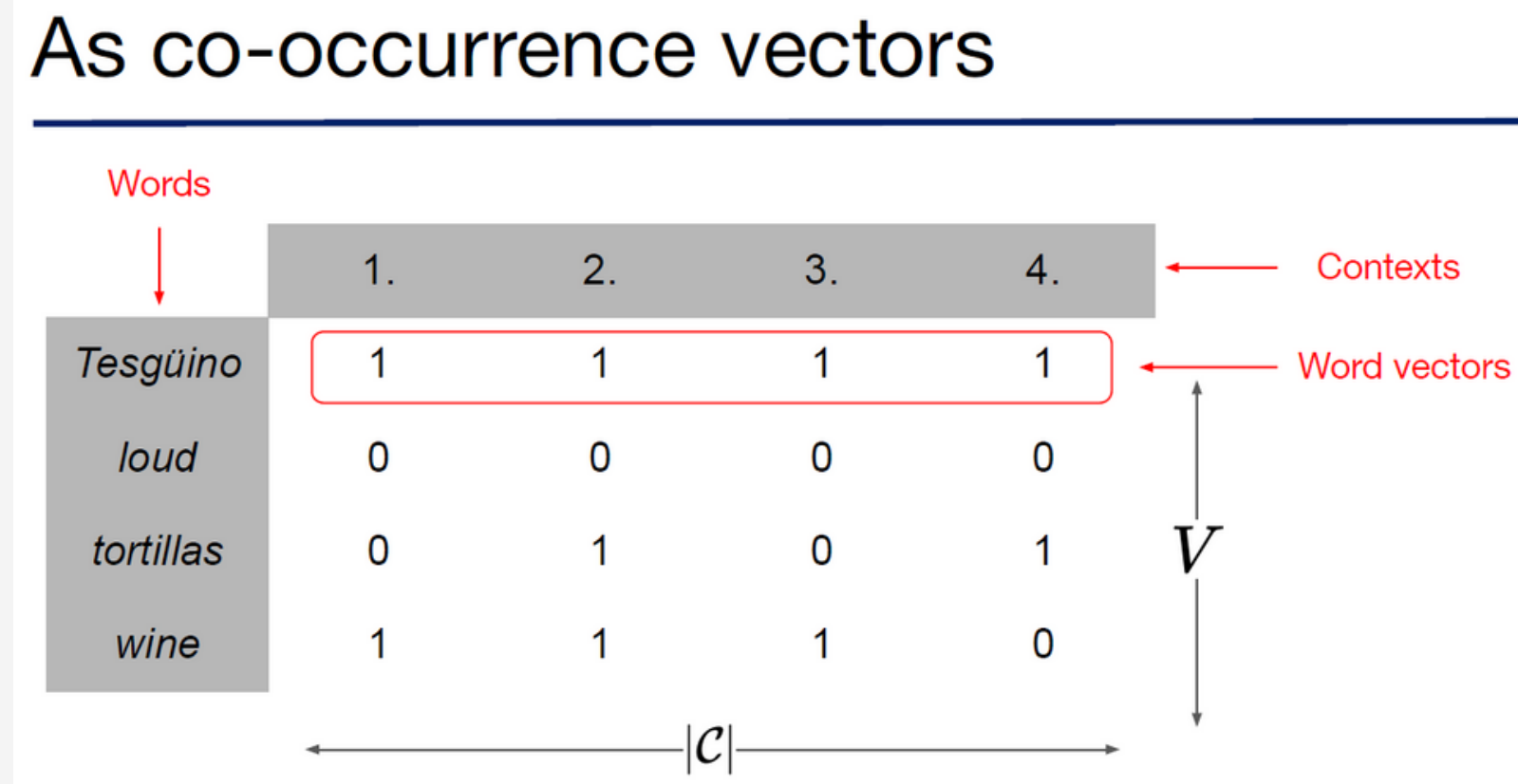
Meaning is use! => need a lot of data

Distributional Hypothesis: Vector similarity implies semantic similarity

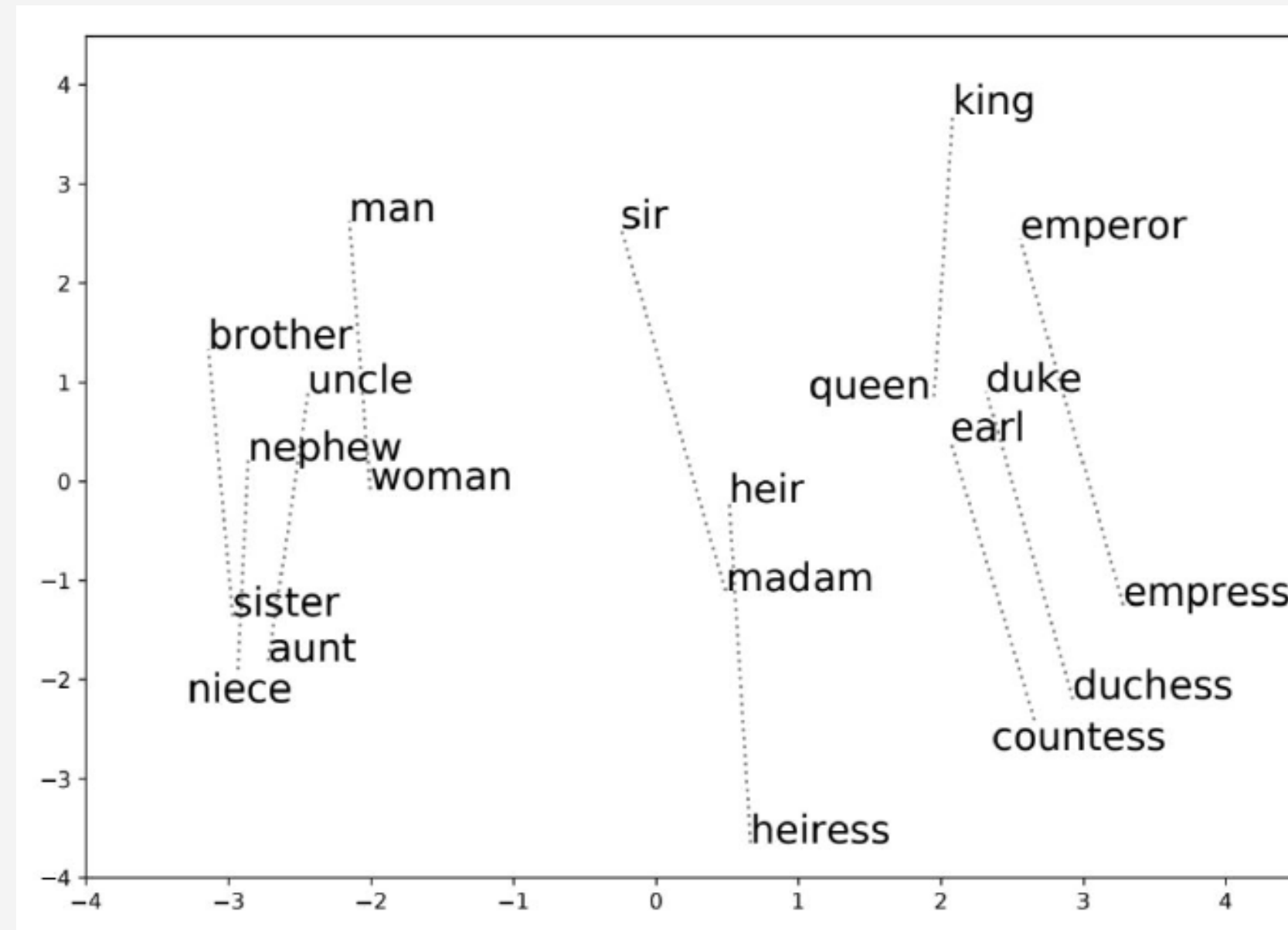
What is the meaning of Tesgüino?

Some contexts:

1. A bottle of Tesgüino is on the table
2. Everybody likes Tesgüino
3. Don't have Tesgüino before your drive
4. We make Tesgüino out of corn



Intrinsic evaluation



- a. Do the embeddings agree with human intuitions of word similarity?
- b. Do the embeddings preserve word analogies?

PIPELINE

March 24, 2023

identify
polysemous words

identify key
philosophical concepts



represent
these words
summarize
themes

word embedding
(skipgram - word2vec)

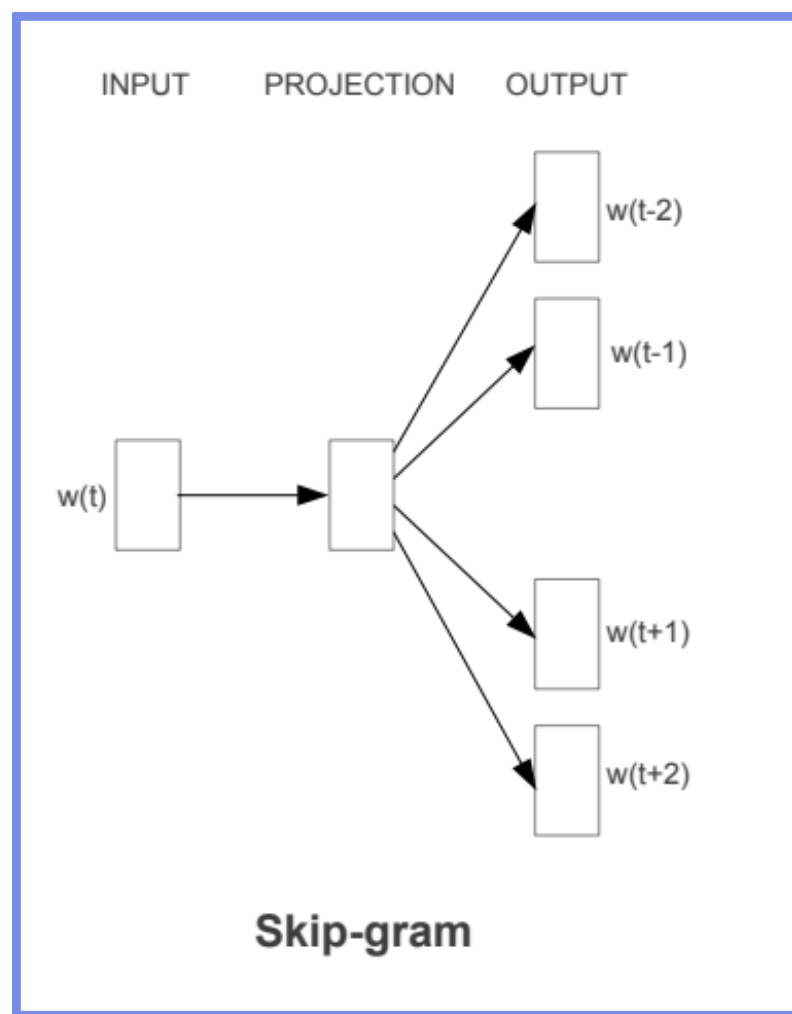
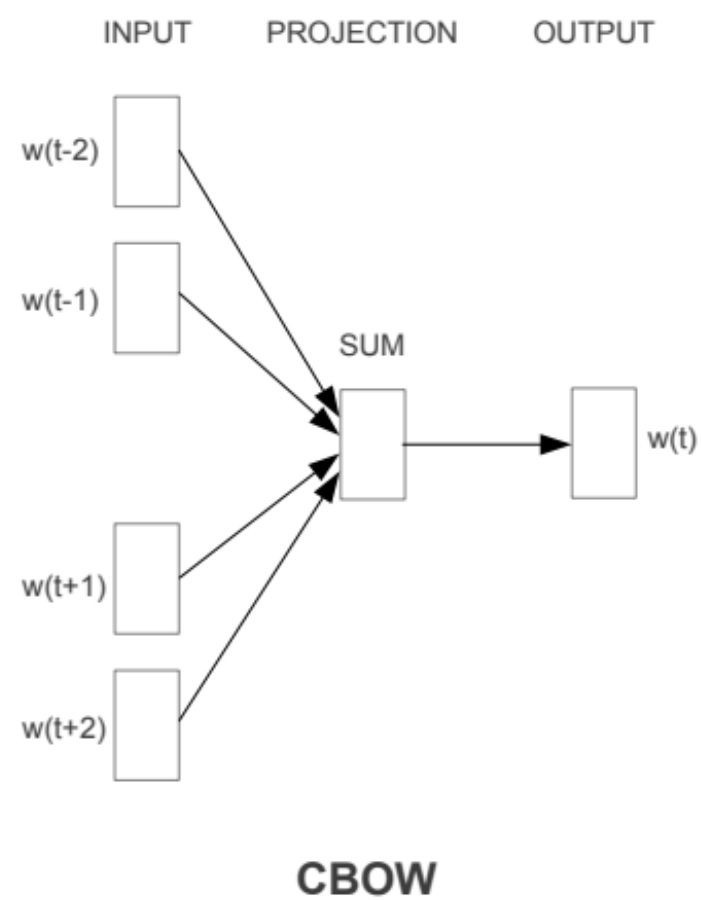
meanings
(in each schools)

interpret features
of schools



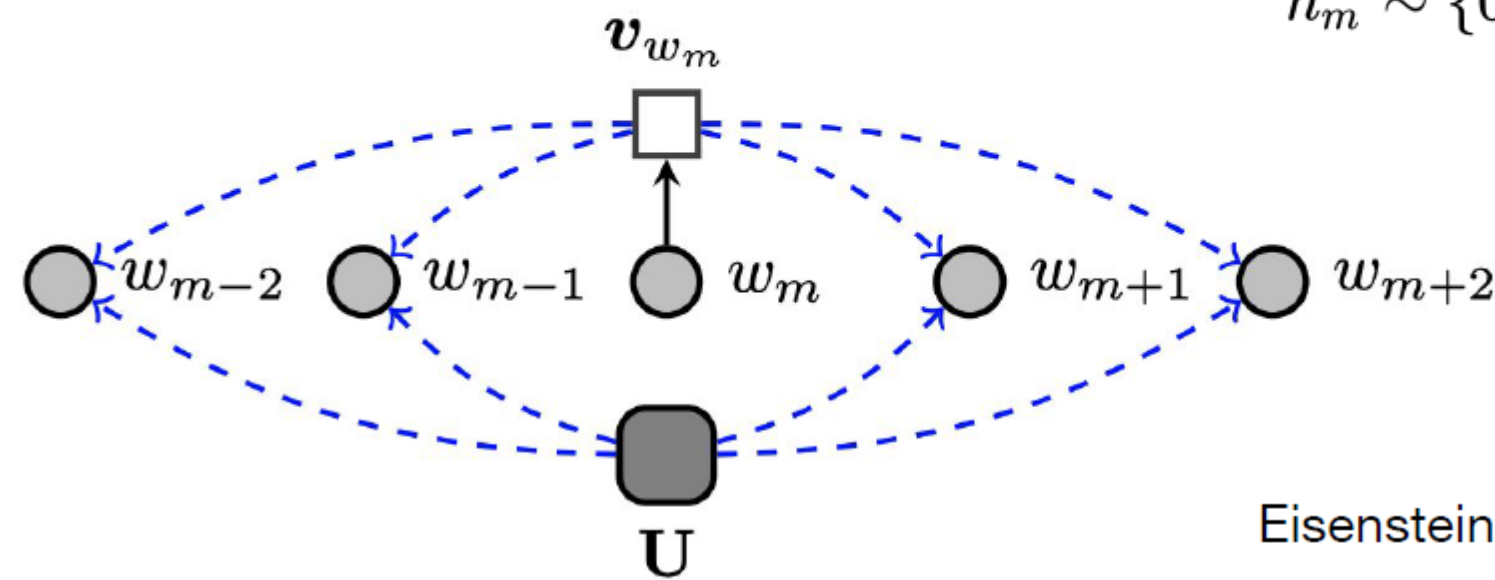
try: context embedding (bert)

Methods: *Word2vec (Skipgram)*



$$\log p(\mathbf{w}) \approx \sum_{m=1}^M \sum_{n=1}^{h_m} \log p(w_{m-n} | w_m) + \log p(w_{m+n} | w_m)$$

$$h_m \sim \{0, \dots, H_{max}\}$$



Eisenstein, Figure 14.3

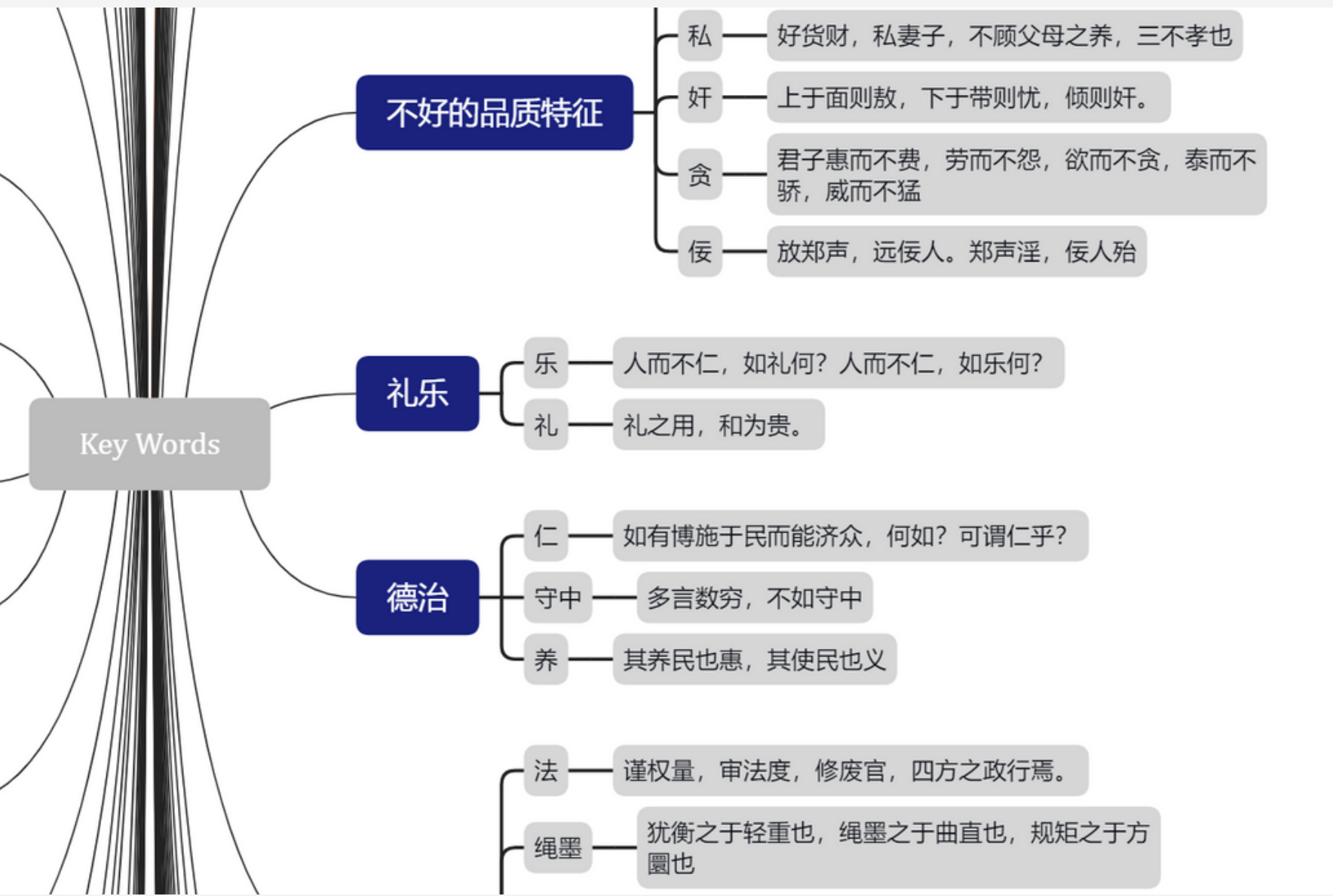
Figure 1: New model architectures. The CBOW architecture predicts the current word based on the context, and the Skip-gram predicts surrounding words given the current word.

Content - close reading

Summarize key concepts and themes

Keywords
-> themes

	A	B	C	D	E
1	School	Primary classification	Secondary classification	Tertiary classification	Key word
2	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	子
3	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	汝
4	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	君
5	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	王
6	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	民
7	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	将
8	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	己
9	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	公
10	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	氏
11	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	母
12	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	神明
13	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	吏
14	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	帝
15	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	卿
16	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	帝王
17	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	寡人
18	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	兵
19	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	三皇
20	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	百姓
21	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	百官
22	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	圣人
23	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	大臣
24	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	至人
25	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	父母
26	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	婴儿
27	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	稚子
28	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	小人
29	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	君子
30	道家Daoism	抽象概念abstract conception	治国理政governance	人称职位title or position	天子



Content - close reading

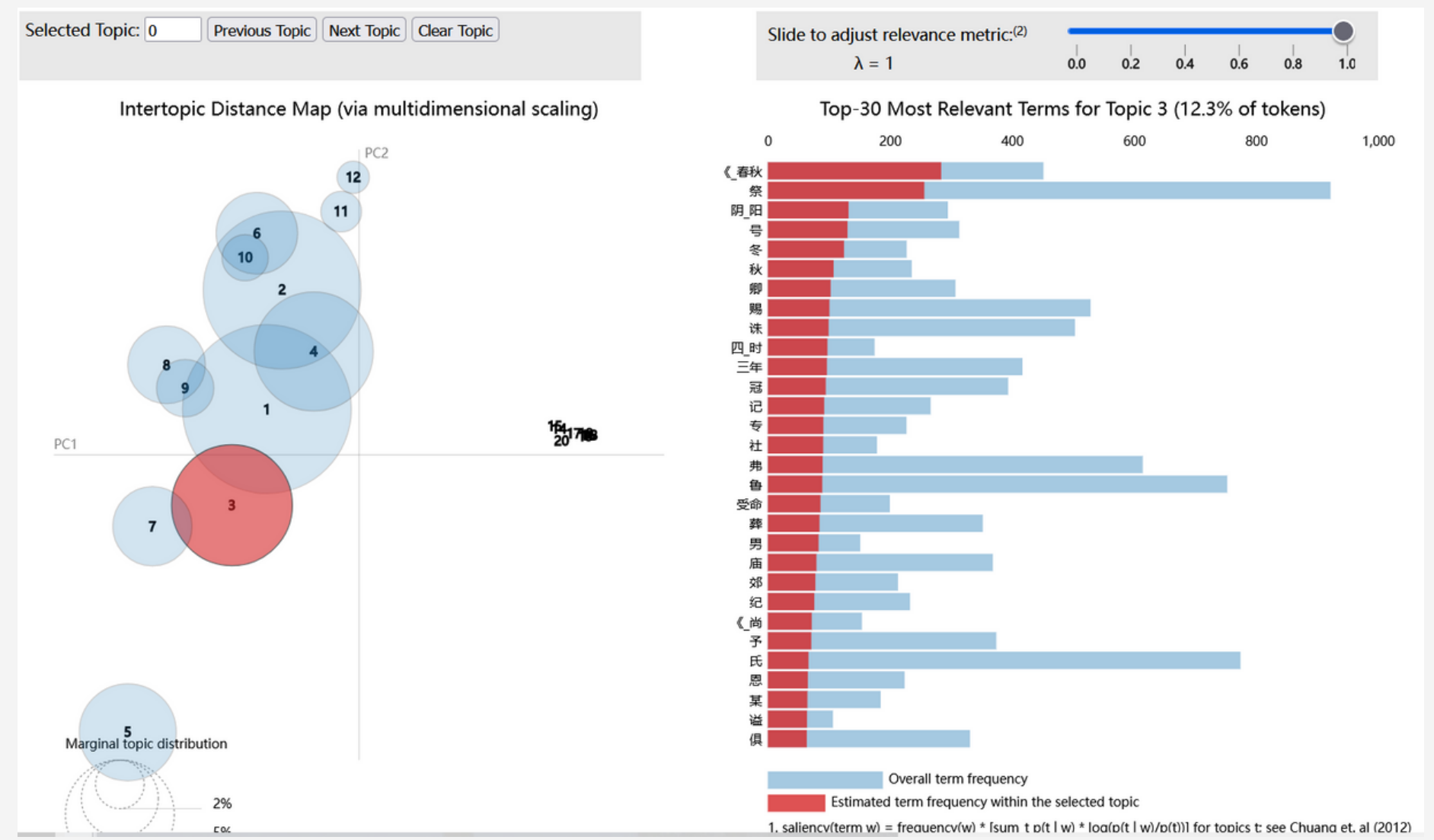
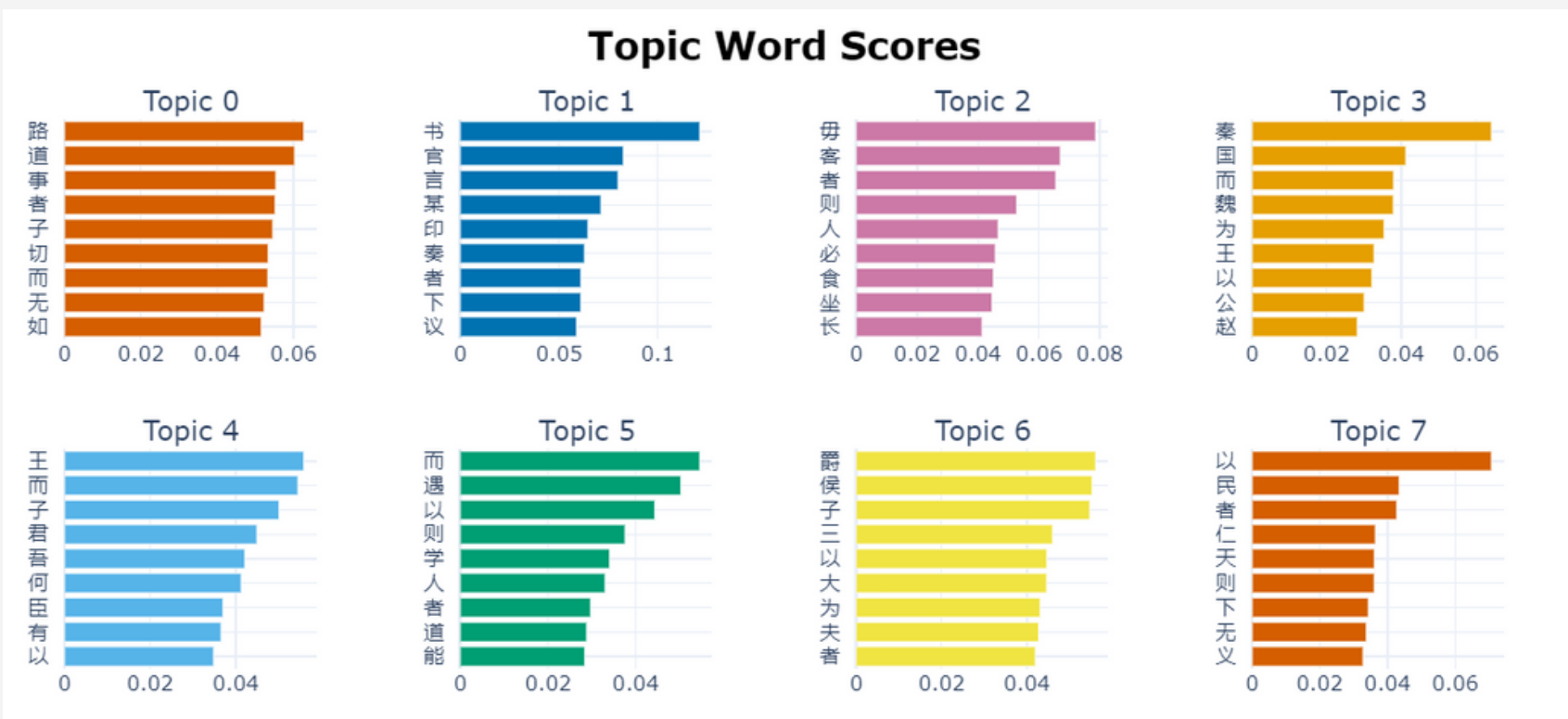
Summarize key concepts and themes

Keywords
-> themes

1	Meaning	Source	school of thought	title	original text	english translation	key concept\	interpretation
2	自然界 material world	https://ctext.org/heshangqong/23/zhs?searchmode=showall#result	道家	《老子河上公章句·虚无》	希言自然。希言者，謂愛言也。愛言者，自然之道。故飄風不終朝，驟雨不終日。飄風，疾風也。驟雨，暴雨也。言疾不能長，暴不能久也。孰為此者？天地。孰，誰也。誰為此飄風暴雨者乎？天地所為。天地尚不能久，不能終於朝暮也。而況於人乎？天地至神合為飄風暴雨，尚不能使終朝至暮，何況人欲為暴卒乎。故從事於道者，從，為也。人為事當如道安靜，不當如飄風驟雨也。道者同於道，道者，謂好道人也。同於道者，所謂與道同也。德者同於德，德者，謂好德之人也。同於德者，所謂與德同也。失者同於失。失，謂任己而失人也。同於失者，所謂與失同也。同於道者，道亦樂得之。與道同者，道亦樂得之也。同於德者，德亦樂得之，與德同者，德亦樂得之也。同於生者，生亦樂得之，與生同者，生亦樂得之也。信不		風、雨、德、朝、日、天地、人、从、道、失、同、乐、信、云、龙、风、虎、水、火	
3	自然界 material world	https://ctext.org/huainanzi/yuan-dao-xun/zhs?searchmode=showall#result	杂家	《淮南子·原道训》	夫萍樹根于水，木樹根於土，鳥排虛而飛，獸躡實而走，蛟龍水居，虎豹山處，天地之性也。兩木相摩而然，金火相守而流，員者常轉，竅者主浮，自然之勢也。		萍、树、木、水、土、鸟、飞、兽、走、蛟龙、虎豹、山、天地、性、金、火、员、传、竅、浮、势	
4	自然界 material world	https://ctext.org/lvshichunren/zhs?searchmode=showall#result	杂家	《吕氏春秋·论人》	何謂反諸己也？適耳目，節嗜欲，釋智謀，去巧故，而游意乎無窮之次，事心乎自然之塗，若此則無以害其天矣。無以害其天則知精，知精則知神，知神之謂得一。凡彼萬形，得一後成。故知一，則應物變化，闊大淵深，不可測也。德行昭美，比於日月，不可息也。豪士時之，遠方來賓，不可塞也。意氣宣通，無所束縛，不可收也。故知知一，則復歸於樸，嗜欲易足，取養節薄，不可得也。離世自樂，中情潔白，不可量也。威不能懼，嚴不能恐，不可服也。故知知一，則可動作當務，與時周旋，不可極也。舉錯以數，取與遵理，不可惑也。言無遺者，集肌膚，不可革也。讒人困窮，賢者遂興，不可匿也。故知知一，則若天地然，則何事之不勝，何物之不應？譬之若御者，反諸己，則車輕馬利，致遠復食而不倦。昔上世之亡主，以罪為在人，故日殺僇而不止，以至於亡而不悟。三代之興王，以罪為在己，故日功而不衰，以至於王。		己、适、节、释、去、欲、智谋、巧故、游、意、无穷、事、心、天、害、精、神、知、一、变化、德行、日月、意气、得、量、极、贤者、物、王、罪、功	
4	自然界 material world	https://ctext.org/ghostvalley/zhs?searchmode=showall#result	杂家	《鬼谷子·损兑法灵著》	損兌者，機危之決也。事有適然，物有成敗，機危之動，不可不察。故聖人以無為待有德，言察辭，合於事。兌者，知之也；損者，行之也。損之兌之，物有不可者，聖人不為之辭。故智者不以言失人之言，故辭不煩而心不虛，志不亂而意不邪。當其難易，而後為之謀；因自然之道以為實。圓者不行，方者不止，是謂大功。益之損之，皆為之辭。用分威散勢之權，以見其兌威，其機危乃為之決。故善損兌		損兑、事、适然、物、成敗、危机、动、察、圣人、德、知、行、言、辞、心、志、谋、难、易、功、益、形势	

Content - distant reading Topic modeling

Keywords -> themes



Suggest representative key words for the close reading team to interpret, select, and categorize

Close reading

Distant reading



Pipeline of Content Analysis (iterative)

March 24, 2023

- human-summarized patterns (of key words and meanings of polysemous words)
- topic modeling-produced results -> suggest representative words
- revise human-selected lists with the results + identify topics and categorize words
- input into word embedding
- separate background and target words (separate the same words in different schools), based on frequency and interpretations
- test out word embeddings
- interpret the results and compare with the pre-made hypotheses

Content - distant reading

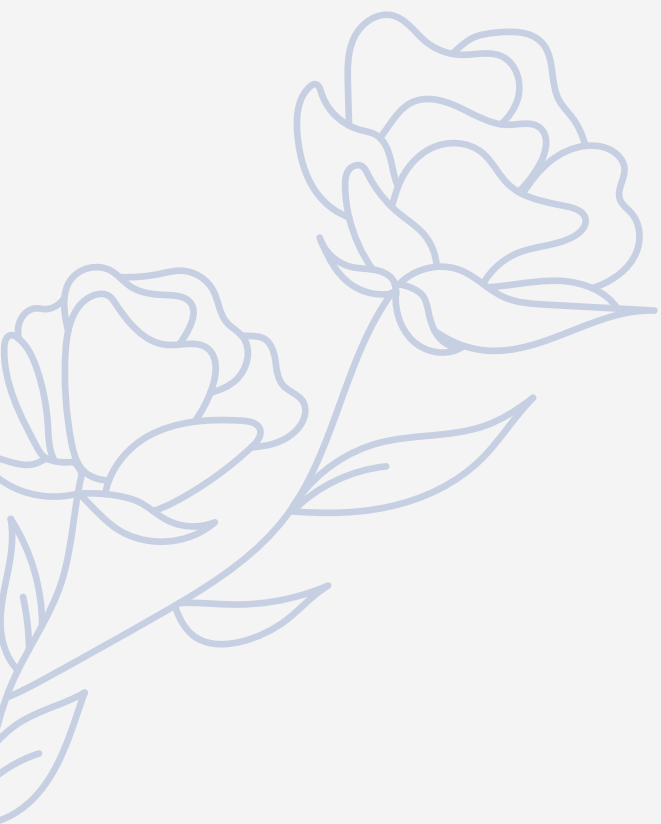
Unsupervised embedding (key concepts)

Hypotheses

1. target words -- for hypothesis testing
 - a. differences in meanings across schools
 - b. relations with other concepts in certain contexts
2. background words -- for pre-defined assumptions

theme	target_word	detailed category	background_word
political/sociological	家	命运	辱
political/sociological	国	命运	宠
political/sociological	民	命运	福
political/sociological	君	命运	中和
person - negative	小人	命运	和气
person - positive	君子	事物变化规律	祸
person - positive	圣人	事物变化规律	变
virtue - positive	仁	事物变化规律	盈
virtue - positive	德	事物变化规律	虚
virtue - positive	美	事物变化规律	妙
virtue - negative	恶	事物变化规律	常
myth/metaphysical	命	事物变化规律	乱
myth/metaphysical	自然	事物变化规律	物
spiritual/intellectual	道	事物变化规律	万物
spiritual/intellectual	理	官职称呼	郡县
spiritual/intellectual	心	官职称呼	寡人
spiritual/intellectual	性	官职称呼	人事
others	法	官职称呼	郡守
others	死	官职称呼	执政
others	病	官职称呼	举兵
others	富贵		

March 24, 2023

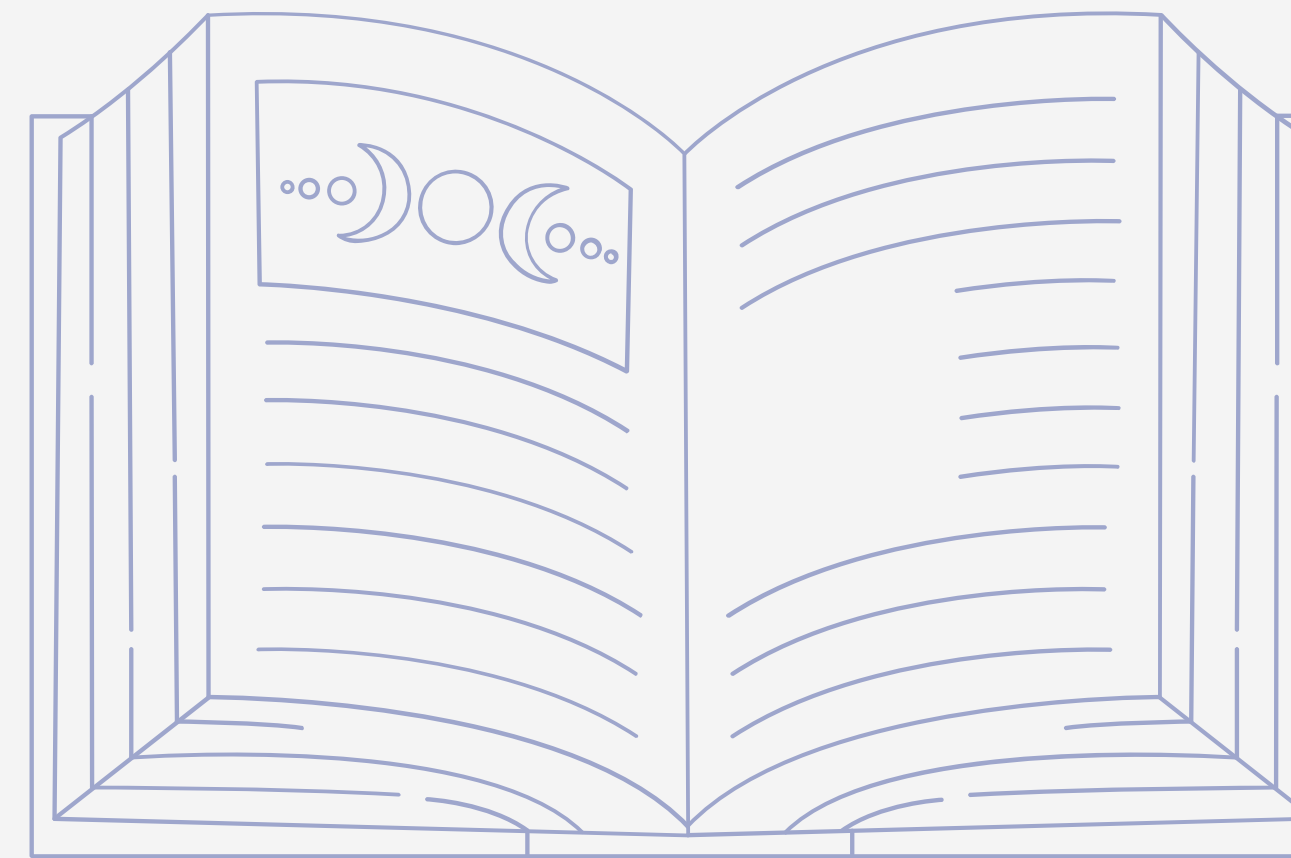


demo

March 24, 2023



Writing Style Analysis



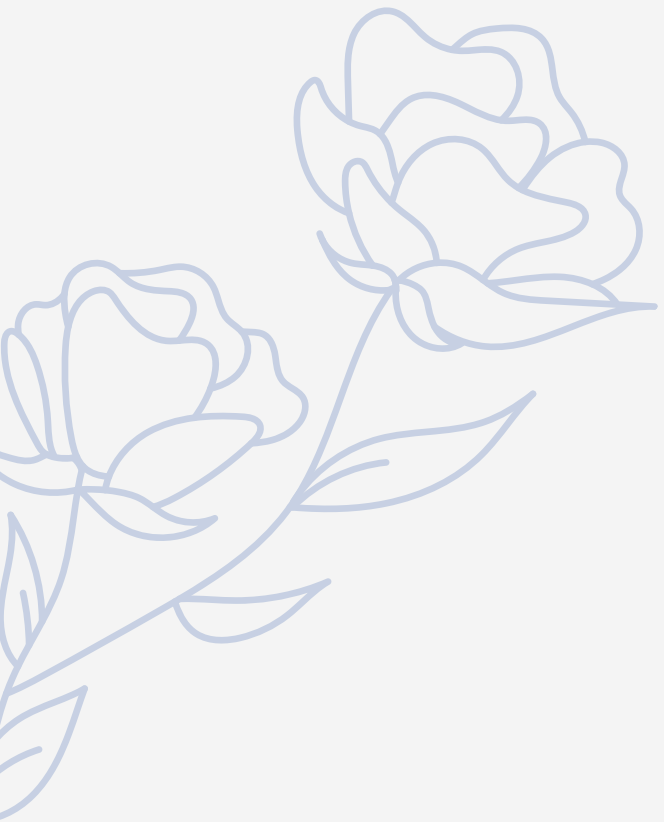
Pipeline of Writing Style Analysis (iterative)

March 24, 2023

- human-summarized patterns (of sentence structures)
- design features for the machine learning model
- analyze feature correlations and importance
- keep experimenting



March 24, 2023



Thank you