

# Visualizing character set collocations *in ancient Chinese literature*

—•••••  
Paul Bucci

# What is CERC?

Cultural Evolution of Religion Research Consortium,  
under the Centre for Human Evolution, Cognition and Culture

## **Extending Evolutionary Accounts of Religion beyond the Mind: Religions as Adaptive Systems**

Purzycki, B. G., Haque, O., and Sosis, R. (2014). Extending Evolutionary Accounts of Religion beyond the Mind: Religions as Adaptive Systems. In Watts, Fraser, and Turner, Léon (eds.). *Evolution, Religion, and Cognitive Science: Critical and Constructive Essays*. Oxford: Oxford University Press, pp. 74-91.

## **Toward a Science of Science Fiction: Applying Quantitative Methods to Genre Individuation.**

Nichols, R., Lynn, J., and Purzycki, B. G. (2014). Toward a Science of Science Fiction: Applying Quantitative Methods to Genre Individuation. *Scientific Study of Literature*, 4(1): 25-45.

## **Decoding a World of Religion – CERC in UC Observer**



They are among the most vexing philosophical questions to confront anyone who has given serious thought to the role of religion in society: Does religious faith, on balance, make humans more moral in their everyday lives? Do some religions promote co-operation better than others? And how did relatively young faiths like Christianity and Islam each [...]

## **Interview with Dr. Edward Slingerland in Religio Magazine**

## Corpus

A set of ancient Chinese texts organized by genre



## Character classes

Sets of characters organized by concept

Gods = {天, 帝 . . . }

Reward = {償, 勞, 胙, 賞, 賜, 酬 . . . }

...

provided by the Cultural Evolution of  
Religion Research Consortium (CERC) lab.

## Task

Identify characters in text  
from each class and calculate  
distances between them

Gods = {天, 帝 . . . }

Reward = {償, 勞, 胙, 賞, 賜, 酬 . . . }

JIAOSHIYILIN

乾道陟石阪胡言連蹇。譯瘖且聾莫使  
道通。請謁不行求事無功。坤招殃來  
螫害我邦國病在手足不得安息。屯陽  
孤亢極多所恨惑車傾蓋亡身常憂惶乃  
得其願雌雄相從。蒙鵠鷓鴣鳩專一無  
尤。君子是則長受嘉福。需目矚足動  
喜如其願舉家蒙寵。訟罷馬上山絕無  
水泉。喉焦唇乾舌不能言。師倉盈庾  
憶宜稼黍稷。國家富有人民蕃息。  
比中夜狗吠盜在牆外。神明祐助銷散  
皆去。小畜据斗運樞順天無憂。與樂  
並居。乾道陟石阪胡言連蹇。譯瘖且

## Gods

### Task

Identify characters in text from each class and calculate distances between them

Gods = {天, 帝 . . . }

Reward = {償, 勞, 胙, 賞, 賜, 酬 . . . }

JIAOSHIYILIN

乾道陟石阪胡言連蹇。譯瘖且聾莫使道通。請謁不行求事無功。坤招殃來螫害我邦國病在手足不得安息。屯陽孤亢極多所恨惑車傾蓋亡身常憂惶乃得其願雌雄相從。蒙鵠鷓鴣鳩專一無尤。君子是則長受嘉福。需目矚足動喜如其願舉家蒙寵。訟罷馬上山絕無水泉。喉焦唇乾舌不能言。師倉盈庾憶宜稼黍稷。國家富有人民蕃息。比中夜狗吠盜在牆外。神明祐助銷散皆去。小畜据斗運樞順天無憂。與樂並居。乾道陟石阪胡言連蹇。譯瘖且

■ Gods    ■ Reward

## Task

Identify characters in text  
from each class and calculate  
distances between them

Gods = {天, 帝 . . . }

Reward = {償, 勞, 胙, 賞, 賜, 酬 . . . }

JIAOSHIYILIN

乾道陟石阪胡言連蹇。譯瘖且聾莫使道通。請謁不行求事無功。坤招殃來螫害我邦國病在手足不得安息。屯陽孤亢極多所恨惑車傾蓋亡身常憂惶乃得其願雌雄相從。蒙鵠鷓鴣專一無尤。君子是則長受嘉福。需目矐足動喜如其願舉家蒙寵。訟罷馬上山絕無水泉。喉焦唇乾舌不能言。師倉盈庾憶宜稼黍稷。國家富有人民蕃息。比中夜狗吠盜在牆外。神明祐助銷散皆去。小畜据斗運樞順天無憂。與樂並居。乾道陟石阪胡言連蹇。譯瘖且

■ Gods    ■ Reward

## Task

Identify characters in text from each class and calculate distances between them

Gods = {天, 帝 . . . }

Reward = {償, 勞, 胙, 賞, 賜, 酬 . . . }

JIAOSHIYILIN

乾道陟石阪胡言連蹇。譯瘖且聾莫使道通。請謁不行求事無功。坤招殃來螫害我邦國病在手足不得安息。屯陽孤亢極多所懼惑車傾蓋亡身常憂惶乃得其願將雄相從。戰鵠鵠鳴鳩專一無尤。君子是則長受嘉福。需目矚足動喜如其願舉家蒙寵。訟罷馬上山絕無水泉。喉焦唇乾舌不能言。師倉盈庾憶宜稼黍稷。國家富有人民蕃息。比中夜狗吠盜在牆外。神明祐助銷散皆去。小畜据斗運樞順天無憂。與樂並居。乾道陟石阪胡言連蹇。譯瘖且

# Data generated

A .csv with collocation data for a focal and comparison class

Number of collocations that occur within  $x$  characters

Number of occurrences of focal class elements in text

DATA.CSV

Genre	Text	Focal	Compare	1	...	50	Sentence	Focal Count	Date
Classics	jiaoshiyilin	Gods	Reward	2	...	125	23	413	400 BCE
Classics	jiaoshiyilin	Gods	Punishment	0	...	22	13	89	30 CE
Classics	jiaoshiyilin	Reward	Punishment	5	...	938	165	1278	200 BCE

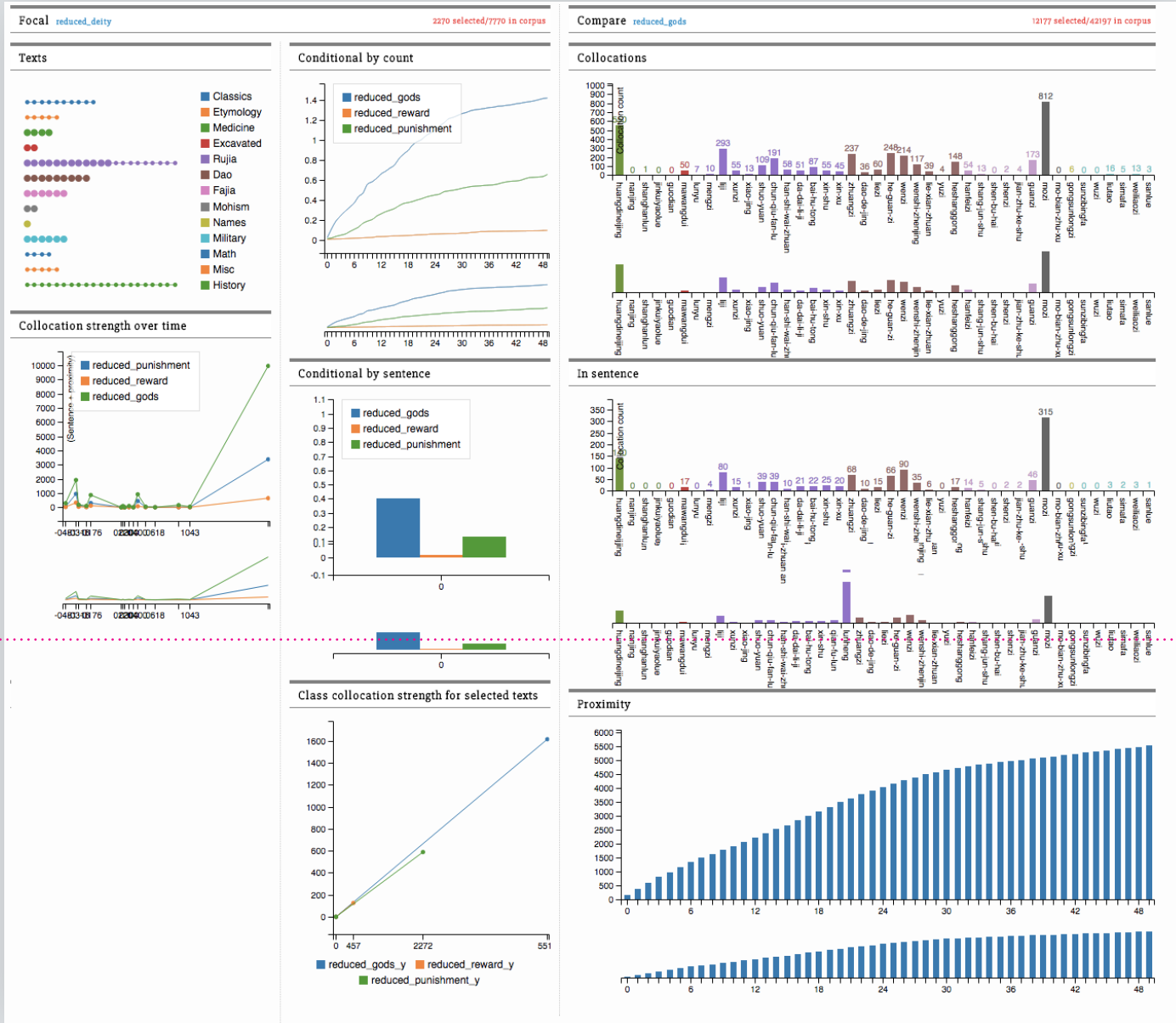
Number of collocations that occur within a sentence



## Interesting statistics

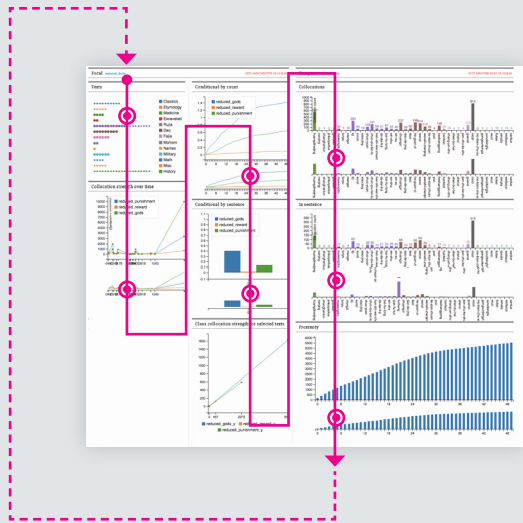
- Frequency of in-sentence collocations.  
*How often do we see these words linked in a sentence?*
- Frequency of collocations within a window of  $k$  characters.  
*How often do we see these words near each other?*
- Conditional probability of focal class to occur within  $k$  characters of comparison class.  
*How likely are two characters to be collocated given the number of occurrences?*
- What happens if we bin the above by genre? Time frame?  
An arbitrary selection of texts?

# Interface overview

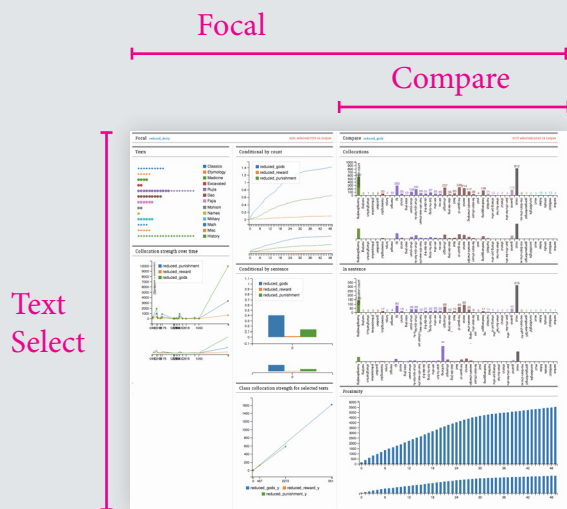


Fold

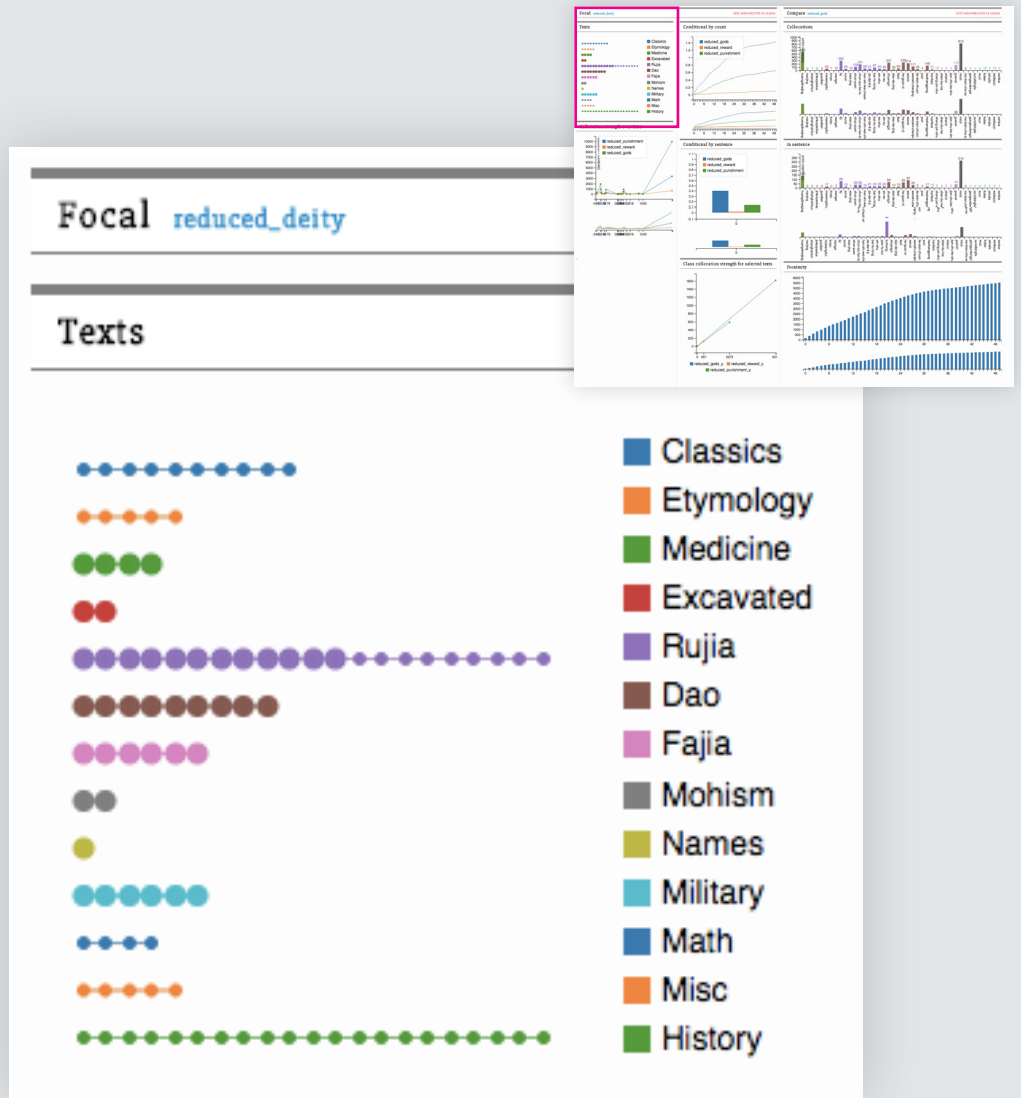
# Workflow



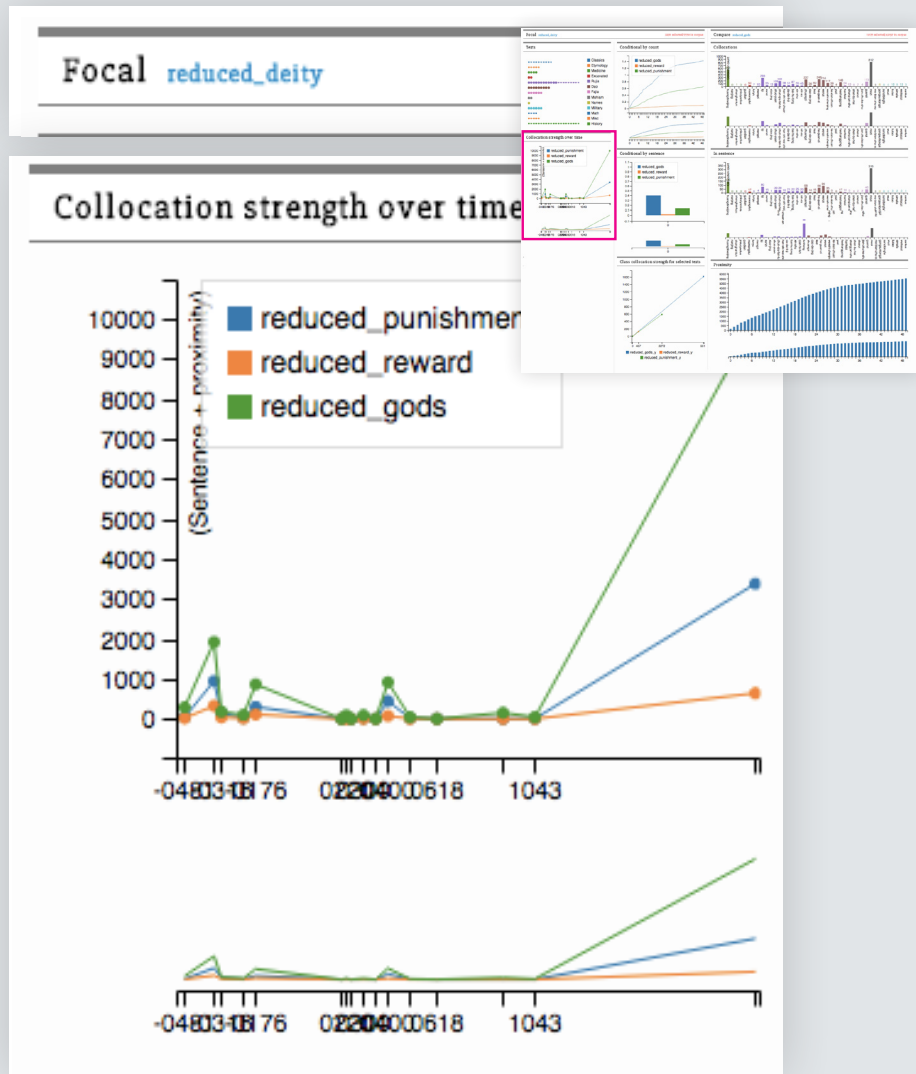
# Influence



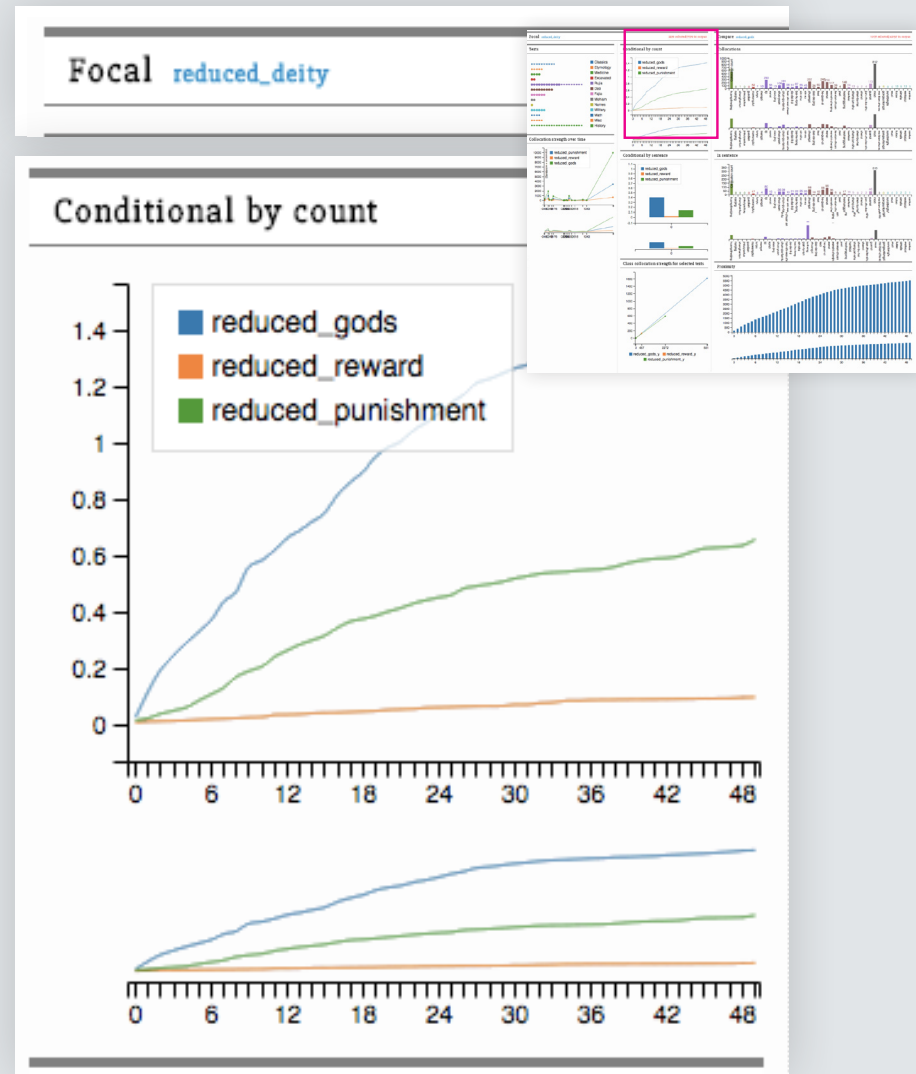
# Text selection



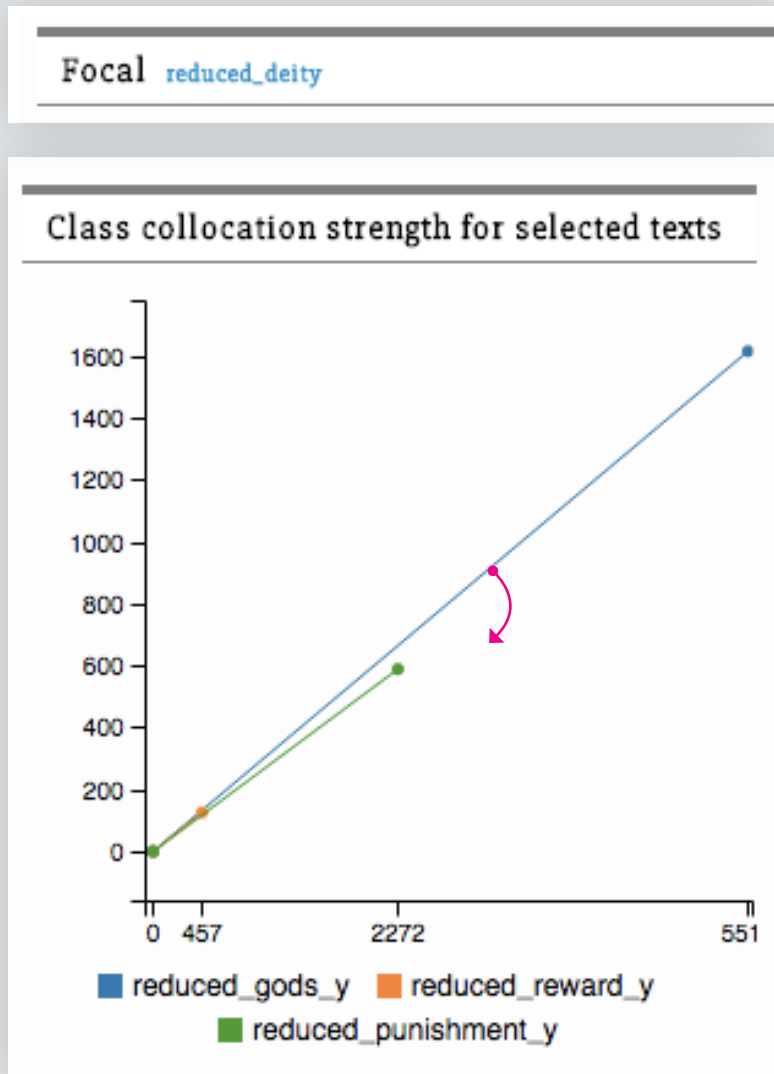
# Collocation strength over time



# Conditional probability evolution and compare



# Strength of collocations between focal and every comparison class



## Three encodings

x-pos = strength by proximity

y-pos = strength by sentence

Angle = ratio difference

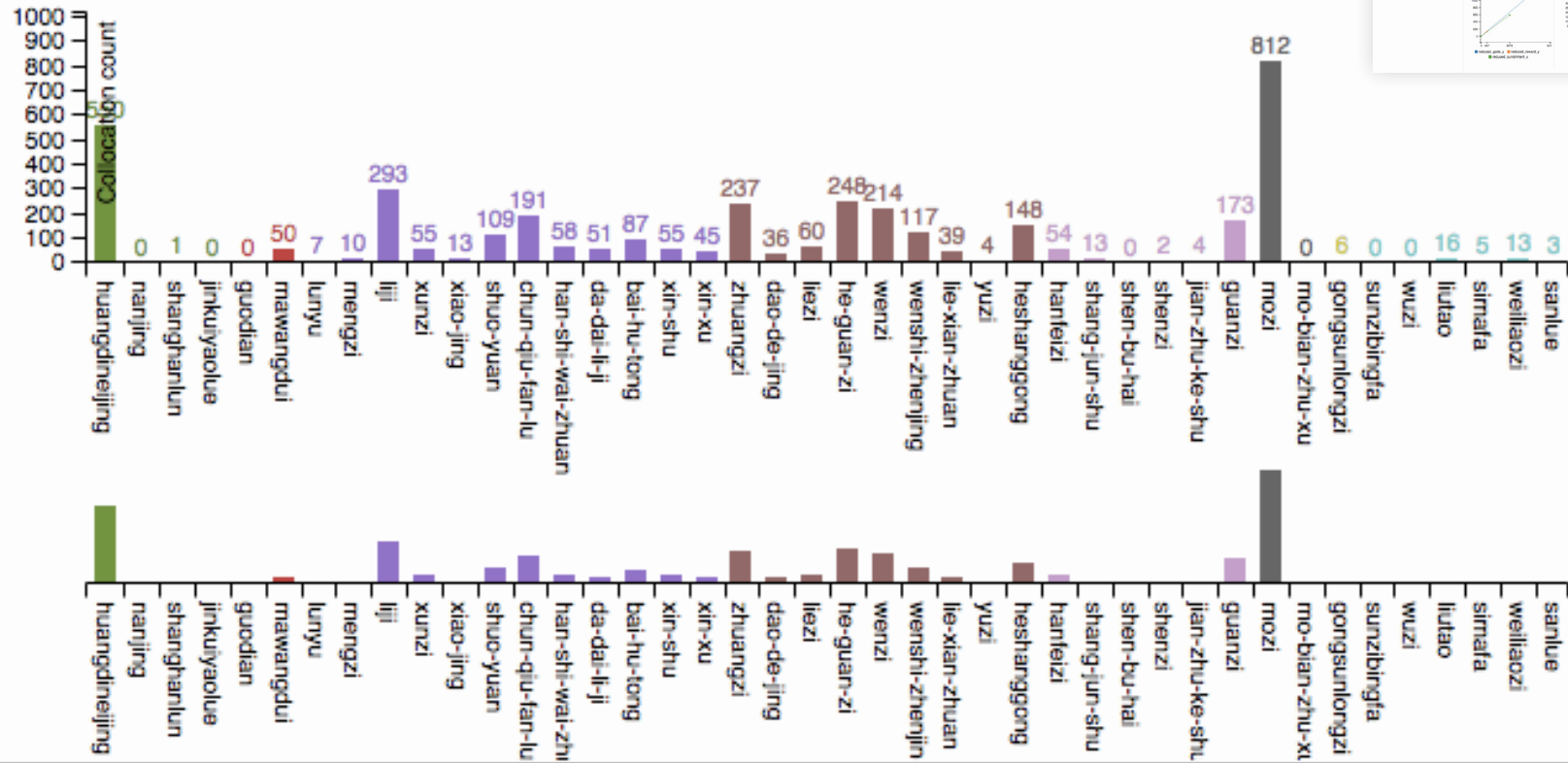


# Collocation counts by text (also in-sentence)

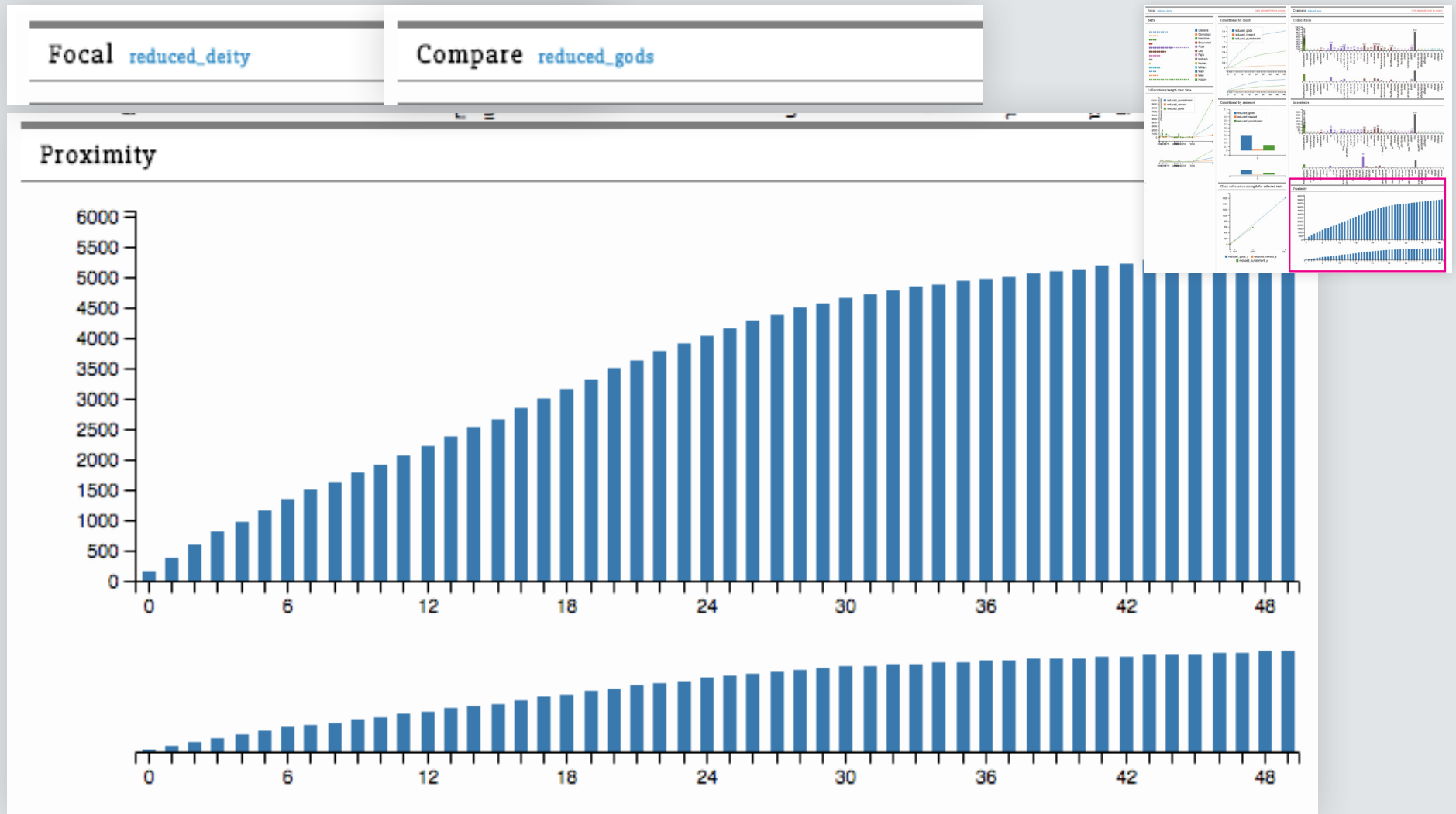
Focal `reduced_deity`

Compare `reduced_gods`

## Collocations



# Collocation counts by proximity



Demo



## Process

- Iteration via in-depth conversations with CERC contact.  
*Involved in every step from planning to evaluation.*
- Some critique for interaction and encoding.  
*Presentation at VIVA lab, discussion.*
- Libraries are amazing and also sometimes terrifying.  
*Had to switch from a Python-only system to D3 in the front and Python in the (far) back.*
- Reception  
*Works for mid-stage questions, proof-of-concept for CERC, seems relatively scalable, shows value of vis for CERC*

## Critique and future plans

- Are these the most elegant encodings?

*Room for a more creative approach, but tradeoff is with ease-of-use.*

- Resources are limited.

*Not enough pixels, but also direction and hierarchy have limitations.*

- Requires a very specific type of CSV.

*But that might be a strength, or at least, an avenue for a strength.*

- Future work

*Scalability, efficiency...back-end integration? Not so sure.*

*Hosting? Customizable probability functions?*

*Generalizability : is there any reason we couldn't do Shakespeare?*

DATA.CSV

Genre	Text	Focal	Compare	1	...	50	Sentence	Focal Count	Date
Classics	jiaoshiyilin	Gods	Reward	2	...	125	23	413	400 BCE
Classics	jiaoshiyilin	Gods	Punishment	0	...	22	13	89	30 CE
Classics	jiaoshiyilin	Punishment	Punishment	5	...	928	165	1278	200 BCE

# Questions?

