



# New ways to read

‘Language on a holiday’

*Machine learning applied to philosophical texts*

Joshua Banks Mailman

jmailman@alumni.uchicago.edu



spaCy

pandas  
 $y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



## Applications



Recommender systems



NLTK

Automated text-to-speech

Education

## **Focus: Distinguishing text that has:**

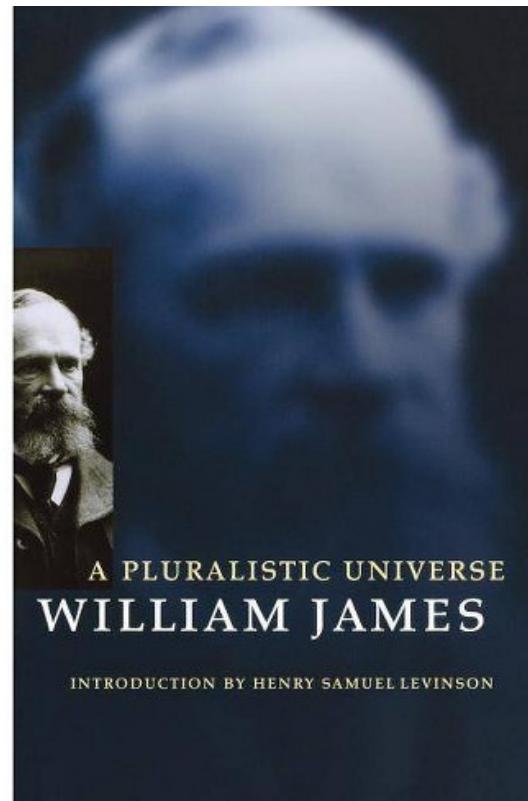
- Topical consistency
- Self-containment (self-explanatory)
- Verbal-semantic eccentricity (idiosyncrasy)

*An*  
**ENQUIRY  
CONCERNING  
HUMAN  
UNDERSTANDING**

**David Hume**



**1748**



**1909**

**Process  
and Reality**

**CORRECTED EDITION**

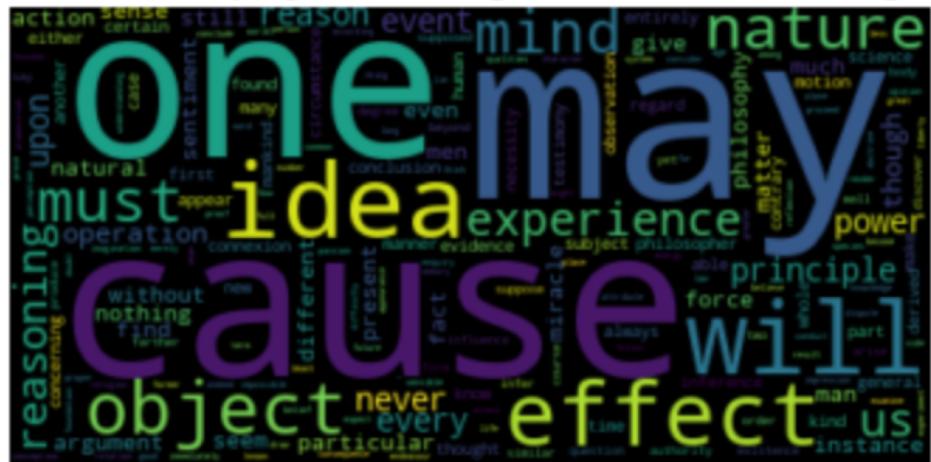


**Alfred North Whitehead**

**Edited by David Ray Griffin  
and Donald W. Sherburne**

**1928**

## Hume's Enquiry Concerning Human Understanding

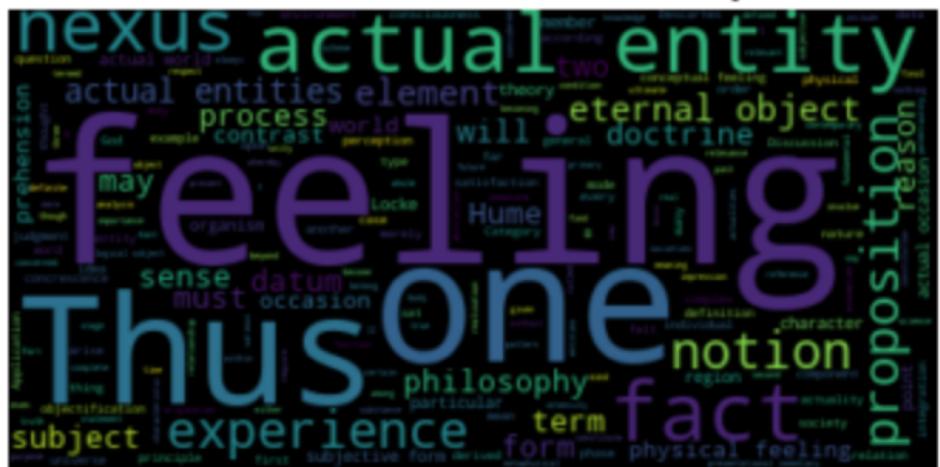


Hume was one of the earliest to draw attention to the distinction: *correlation* vs. *causation*, and issues of observing **cause** and **effect**. Hume is known as among the most lucid philosophers.

James's Pluralist Universe



Whitehead's Process and Reality

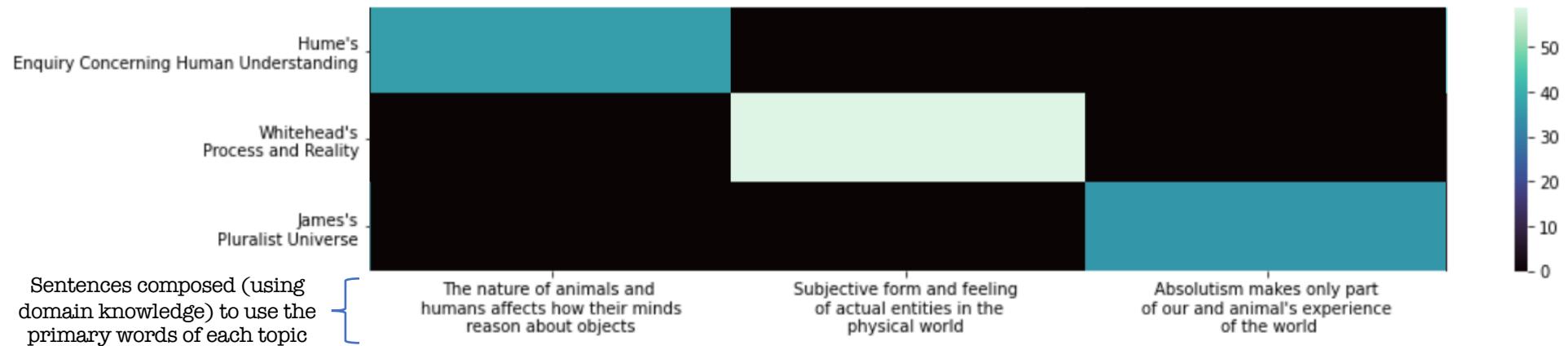


Whitehead, by contrast, is infamous for the obscurity of his ideas and the idiosyncracy of his verbal expression. French philosopher Gilles Deleuze's interest in Whitehead has brought more attention to his work recently.

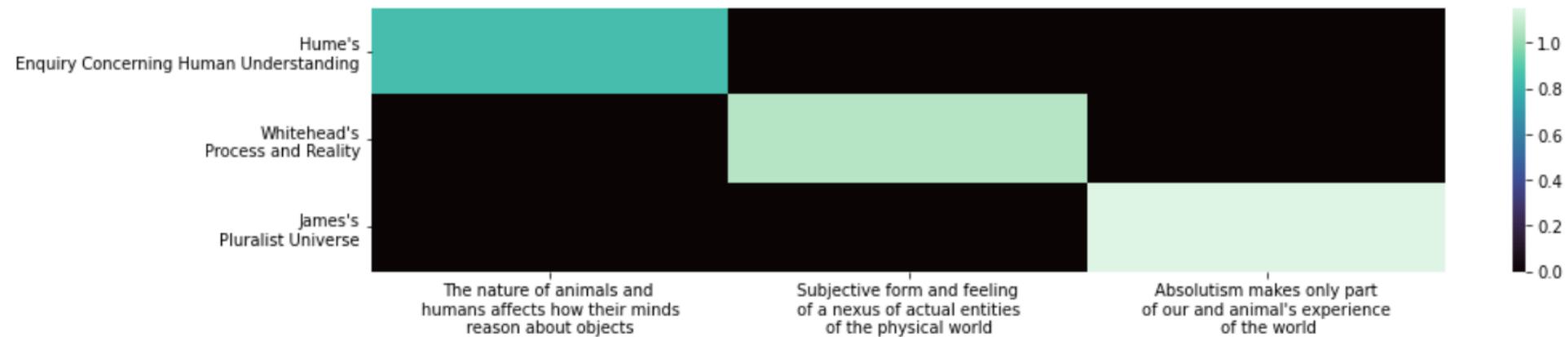
# Topic modeling

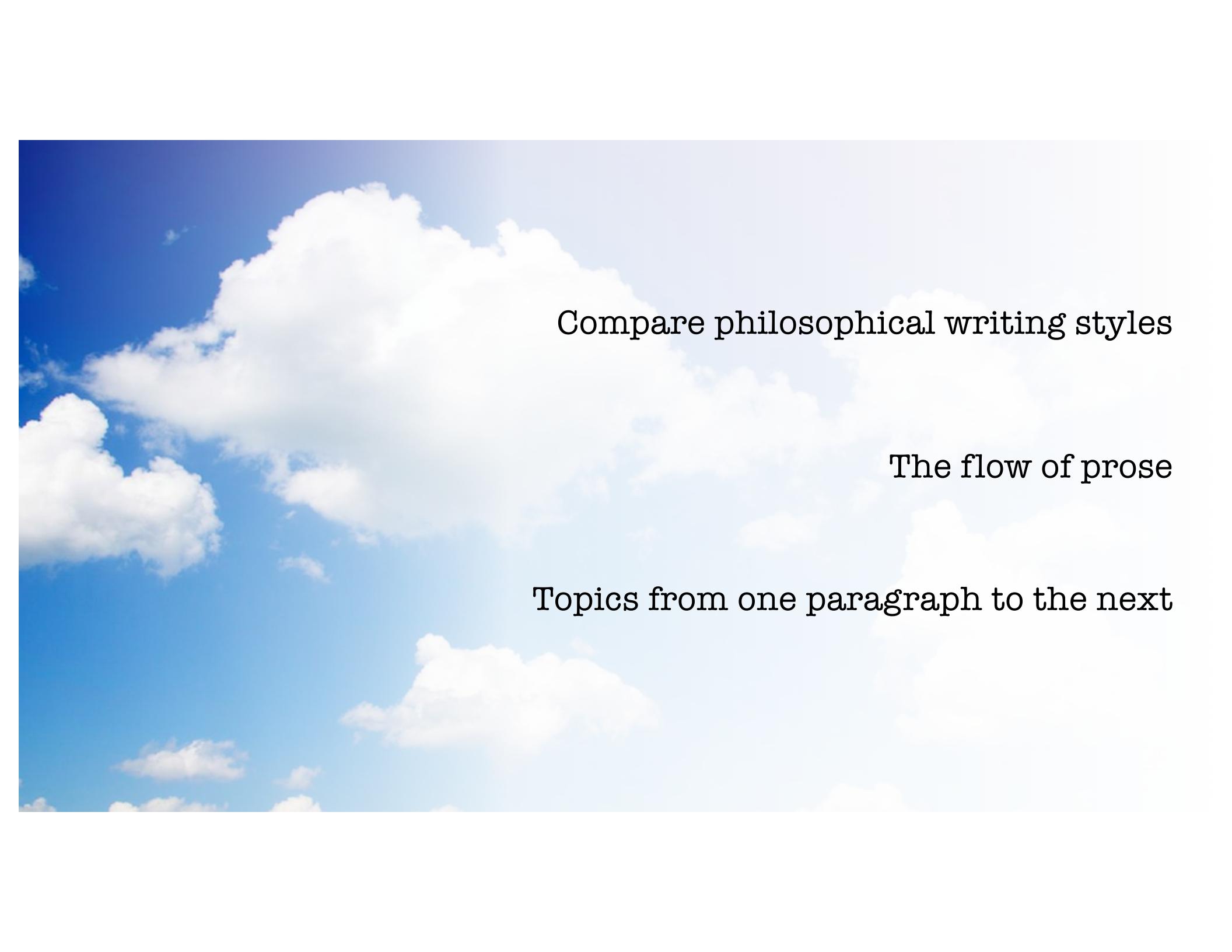
NMF (non-negative matrix factorization)

## snowball stemmer



## word stemmer



The background of the slide is a photograph of a clear blue sky with several large, white, fluffy cumulus clouds scattered across it.

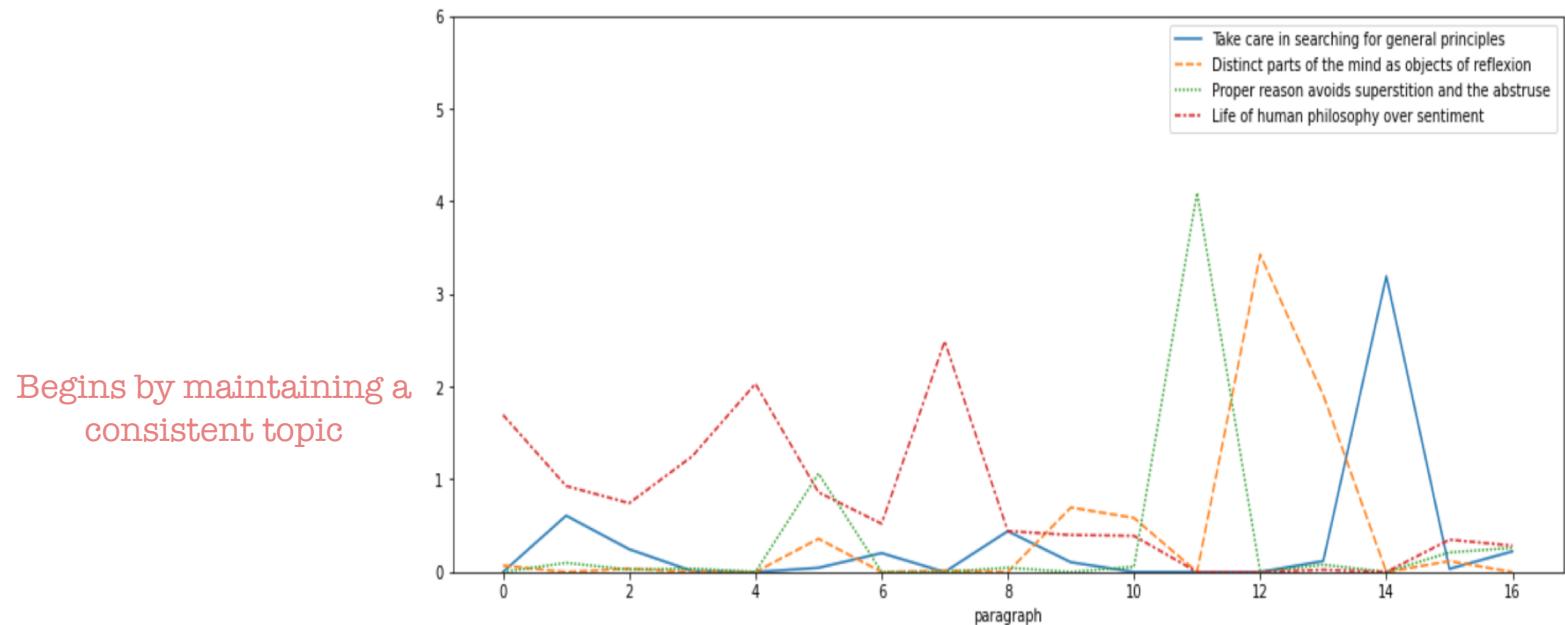
Compare philosophical writing styles

The flow of prose

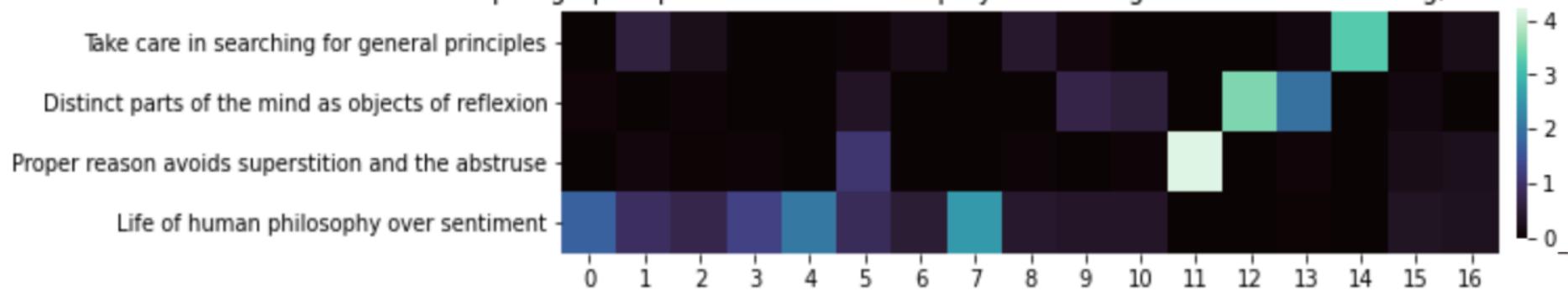
Topics from one paragraph to the next

Hume

## Paragraph topic flow (four topics)

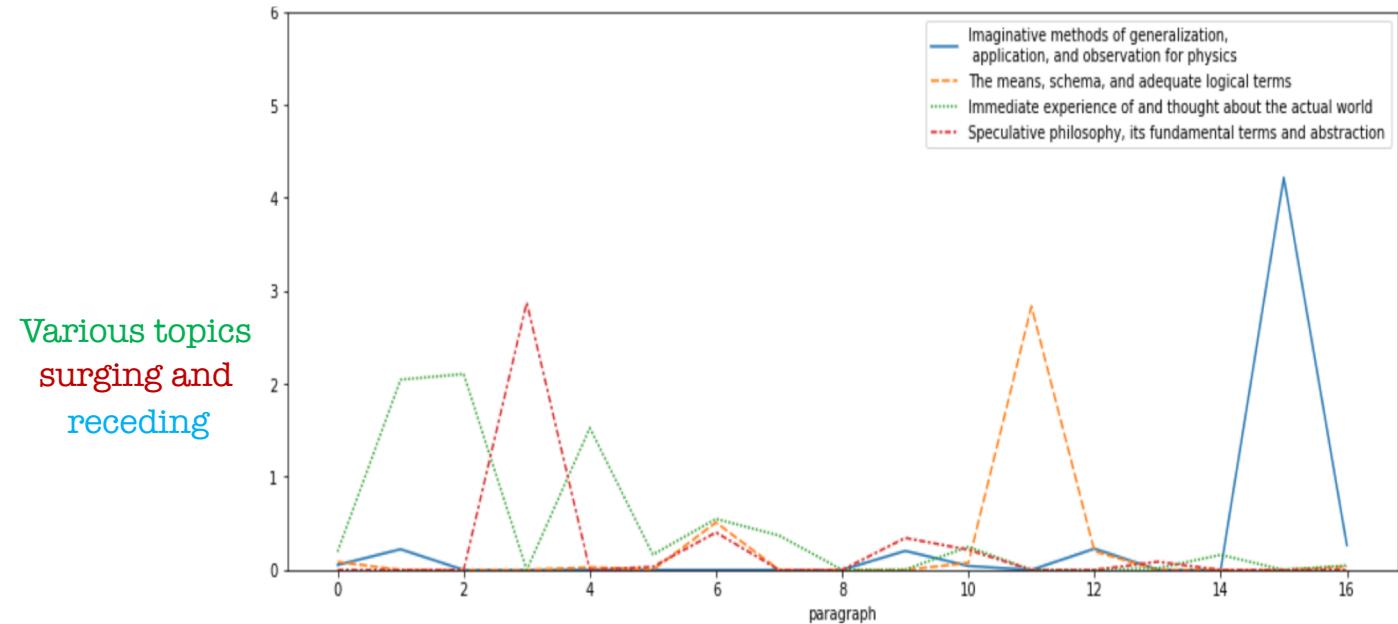


## paragraph topic flow in Hume's Enquiry Concerning Human Understanding, section 1

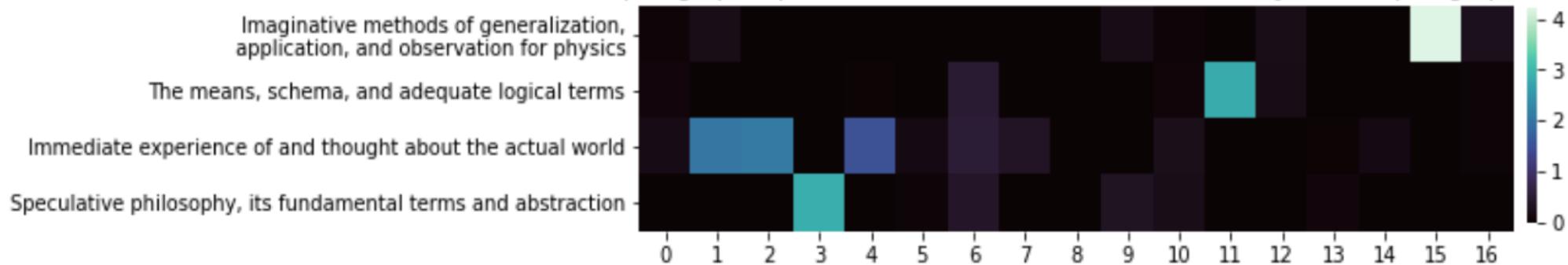


# Whitehead

## Paragraph topic flow (four topics)

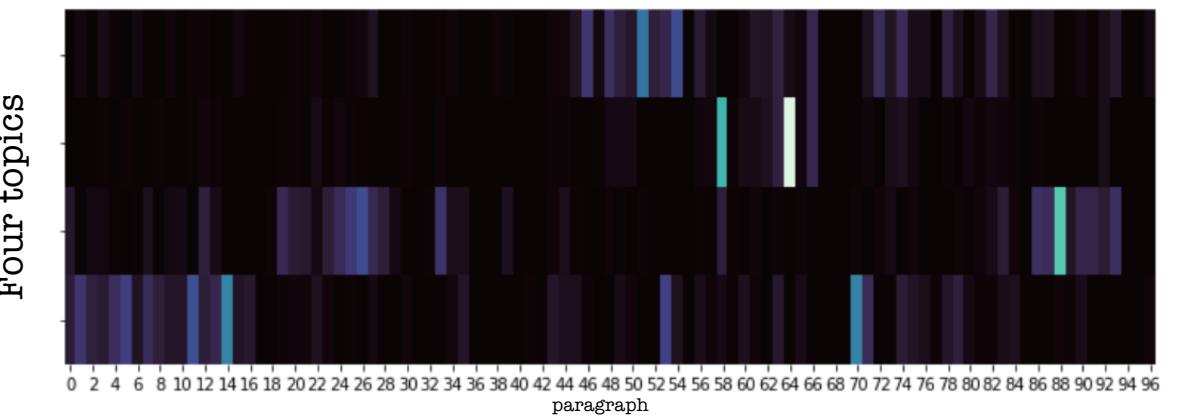


paragraph topic flow in Whitehead's Process and Reality, first 17 paragraphs

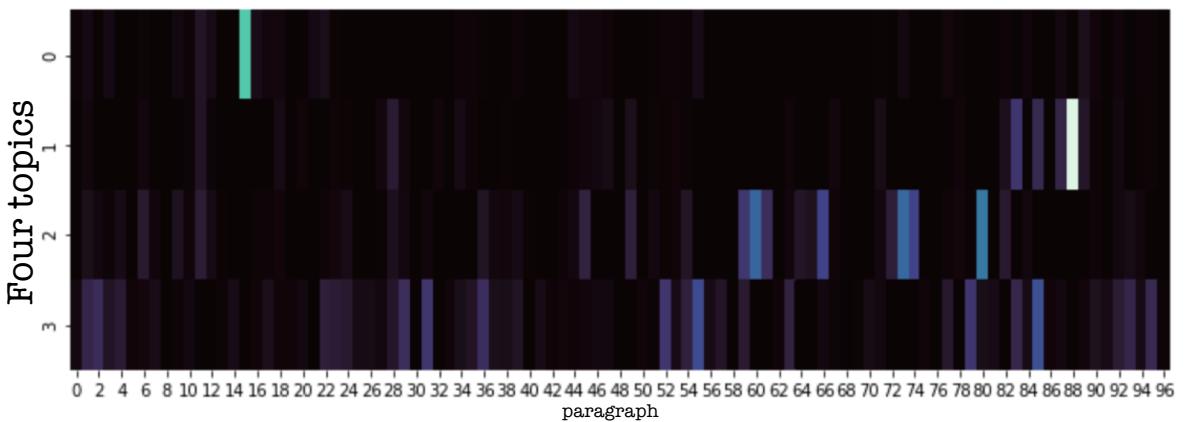


## Paragraph topic flow in longer passages

**Hume** (first 96 paragraphs)



**Whitehead's first chapter** (96 paragraphs)

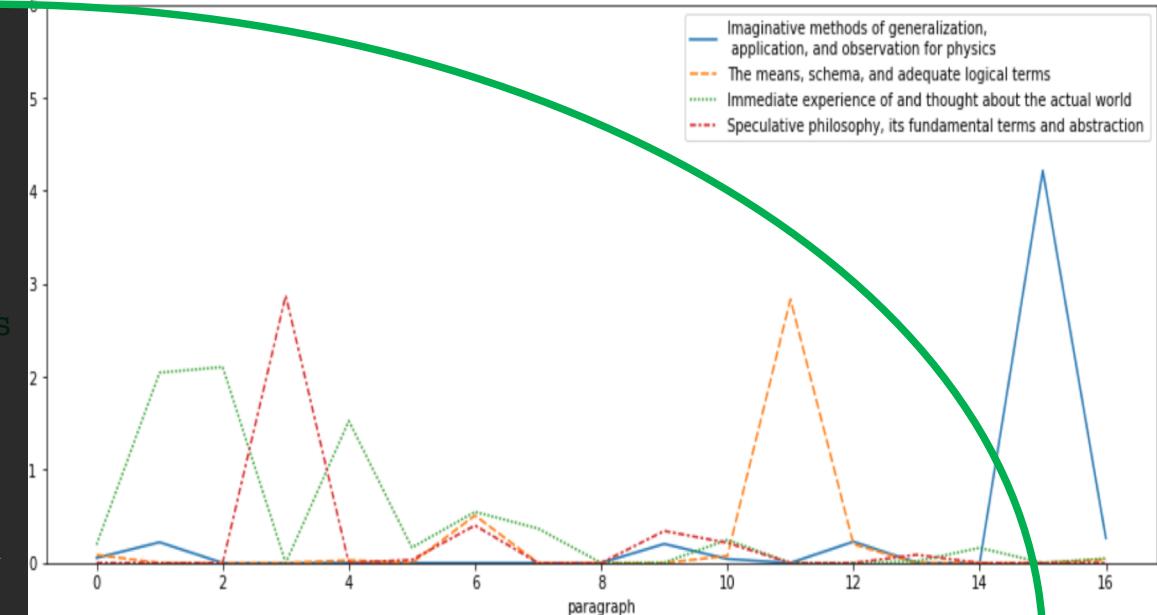


## Whitehead

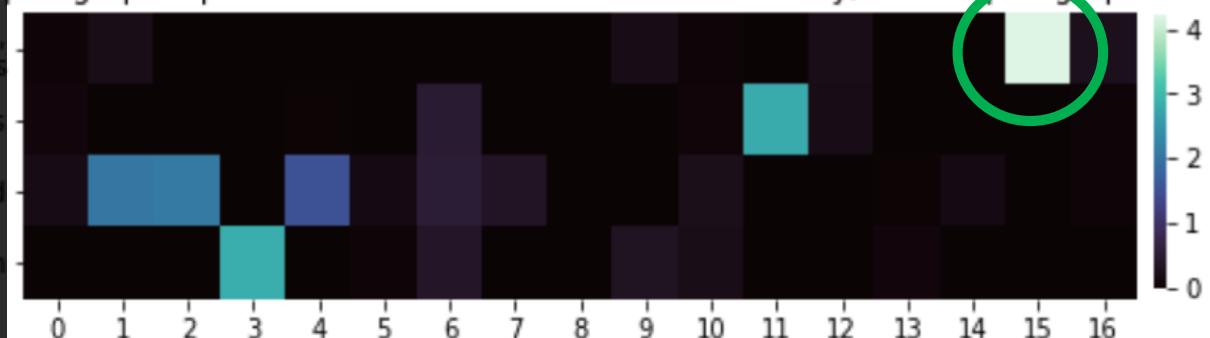
a more free-standing self-explanatory passage

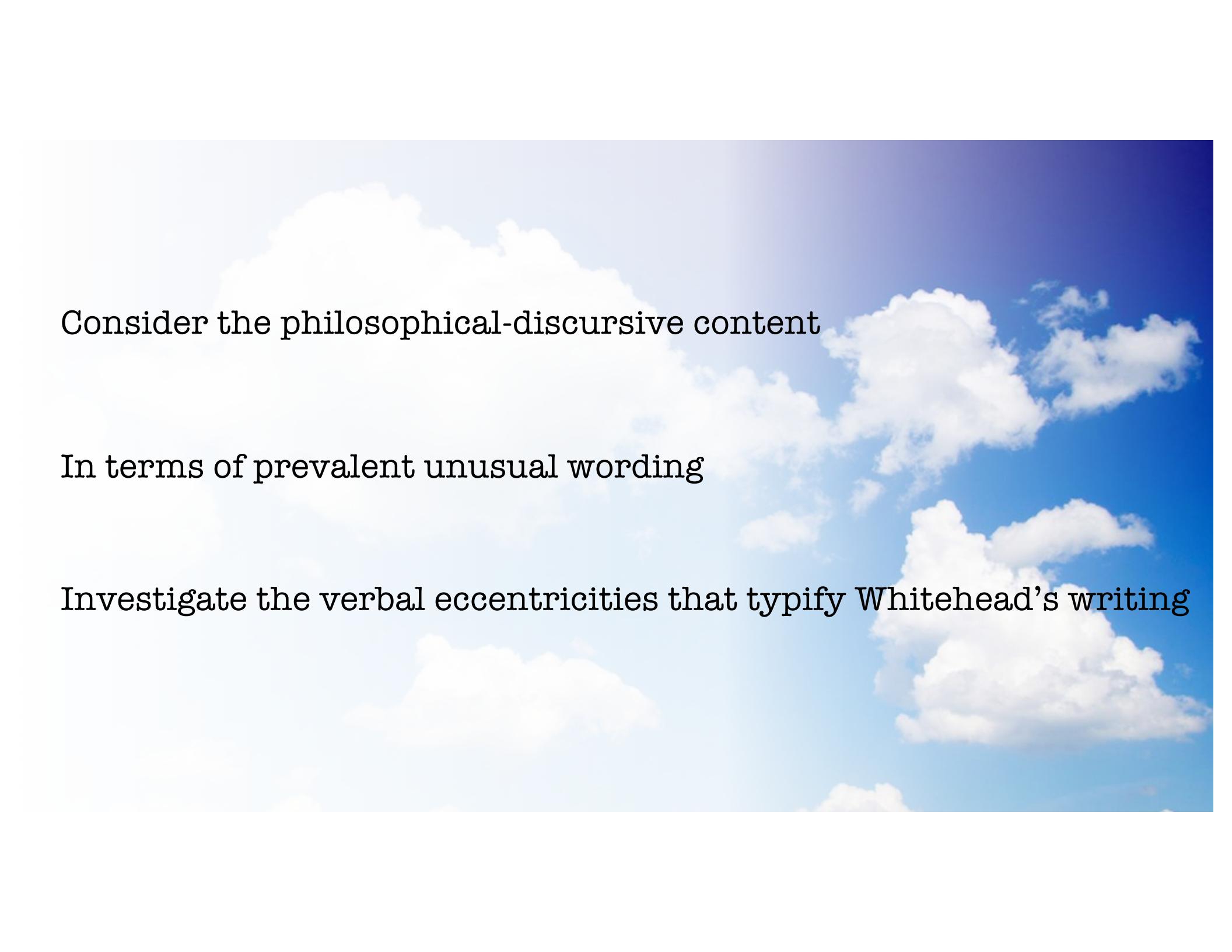
Larger generalities. In natural science this rigid method is the Baconian method of induction, a method which, if consistently pursued, would have left science where it found it. What Bacon omitted was the play of a free imagination, controlled by the requirements of coherence and logic. The true method of discovery is like the flight of an aeroplane. It starts from the ground of particular observation; it makes a flight in the thin air of imaginative generalization; and it again lands for renewed observation rendered acute by rational interpretation. The reason for the success of this method of imaginative rationalization is that, when the method of difference fails, factors which are constantly present may yet be observed under the influence of imaginative thought. Such thought supplies the differences which the direct observation lacks. It can even play with inconsistency; and can thus throw light on the consistent, and persistent, elements in experience by comparison with what in imagination is inconsistent with them. The negative judgment is the peak of mentality. But the conditions for the success of imaginative construction must be rigidly adhered to. In the first place, this construction must have its origin in the generalization of particular factors discerned in particular topics of human interest; for example, in physics, or in physiology, or in psychology, or in aesthetics, or in ethical beliefs, or in sociology, or in languages conceived as storehouses of human experience. In [8] this way the prime requisite, that anyhow there shall be some important application, is secured. The success of the imaginative experiment is always to be tested by the applicability of its results beyond the restricted locus from which it originated. In default of such extended application, a generalization started from physics, for example, remains merely an alternative expression of notions applicable to physics. The partially successful philosophic generalization will, if derived from physics, find applications in fields of experience beyond physics. It will enlighten observation in those remote fields, so that general principles can be discerned as in process of illustration, which in the absence of the imaginative generalization are obscured by their persistent exemplification.

## Paragraph topic flow (four topics)



paragraph topic flow in Whitehead's Process and Reality, first 17 paragraphs

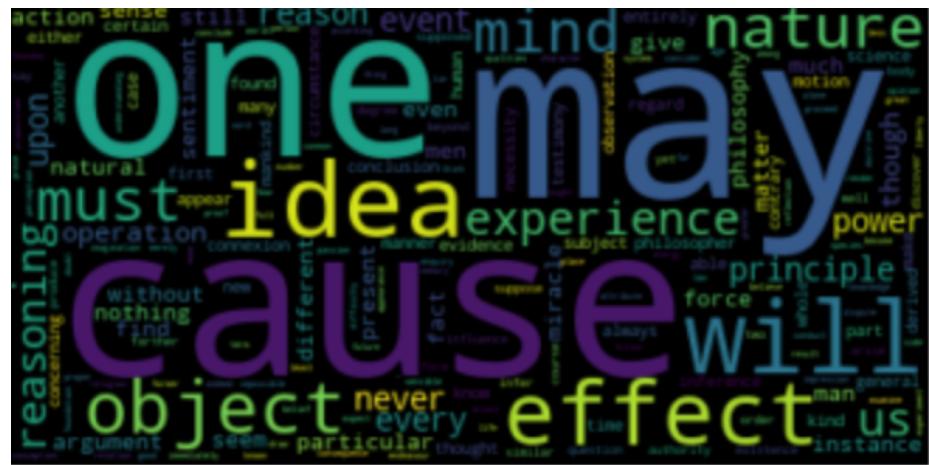


The background of the slide is a photograph of a clear blue sky with scattered white, fluffy clouds.

Consider the philosophical-discursive content

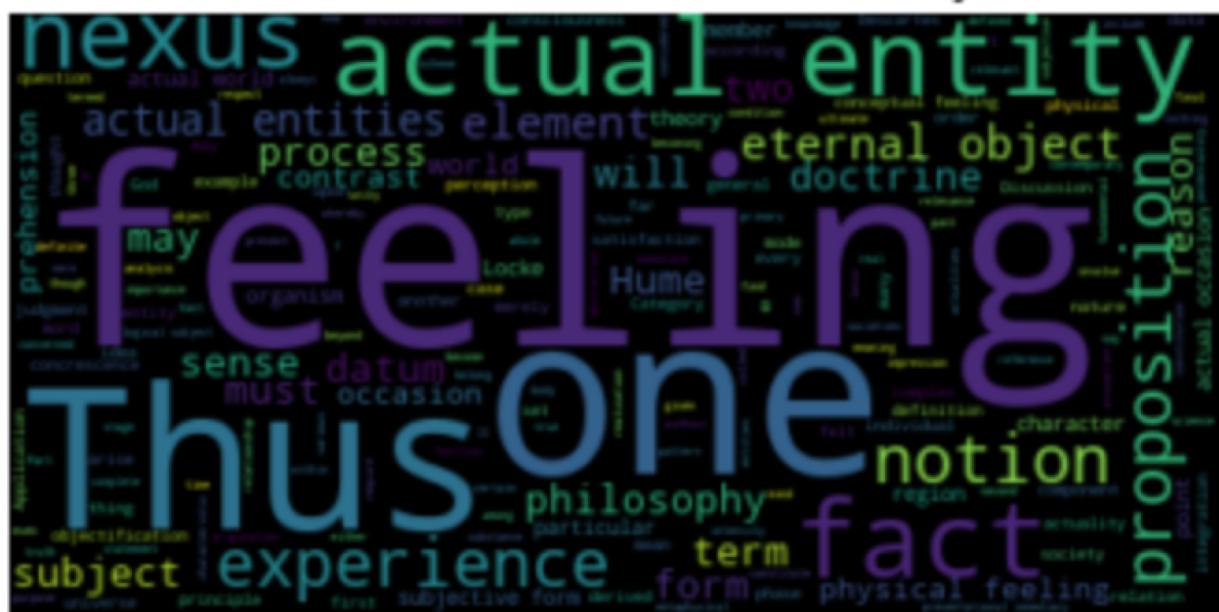
In terms of prevalent unusual wording

Investigate the verbal eccentricities that typify Whitehead's writing

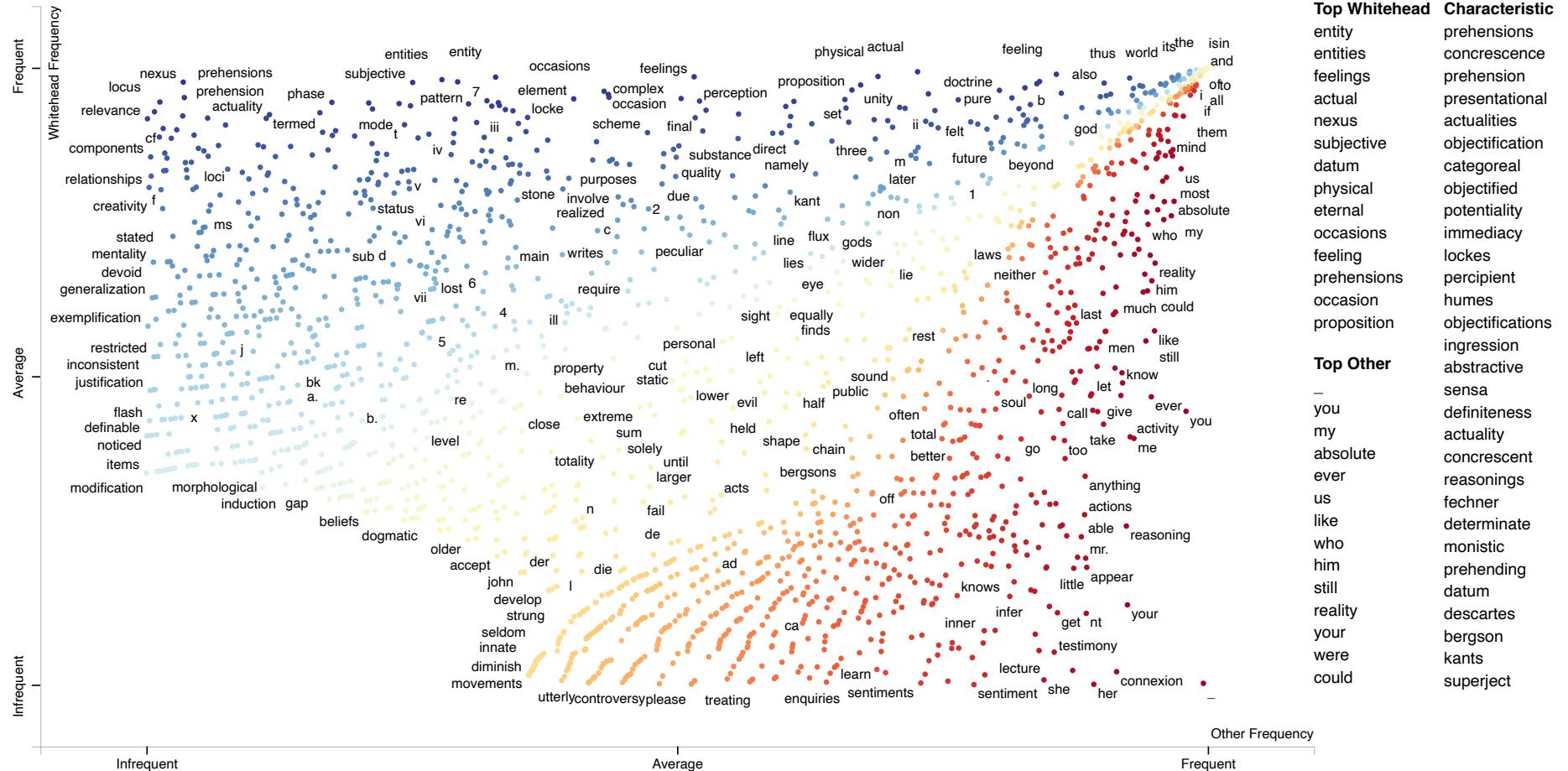


Hume's Enquiry Concerning Human Understanding

## Whitehead's Process and Reality



## Scattertext of Whitehead vs. Hume and James

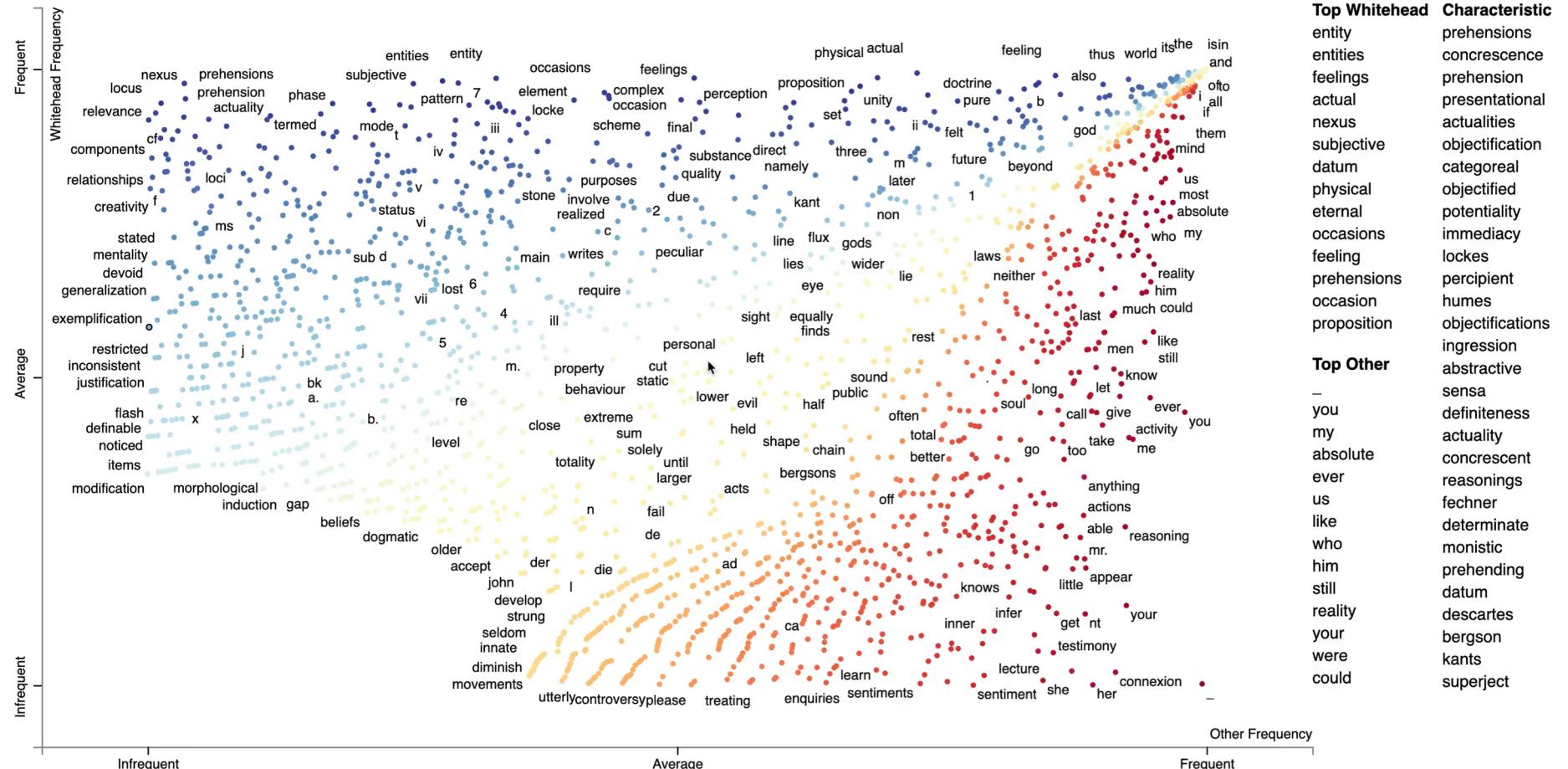


Search the chart

**Whitehead** document count: 1; word count: 157,800  
**Other** document count: 2; word count: 126,615

# spaCy

## Scattertext of Whitehead vs. Hume and James

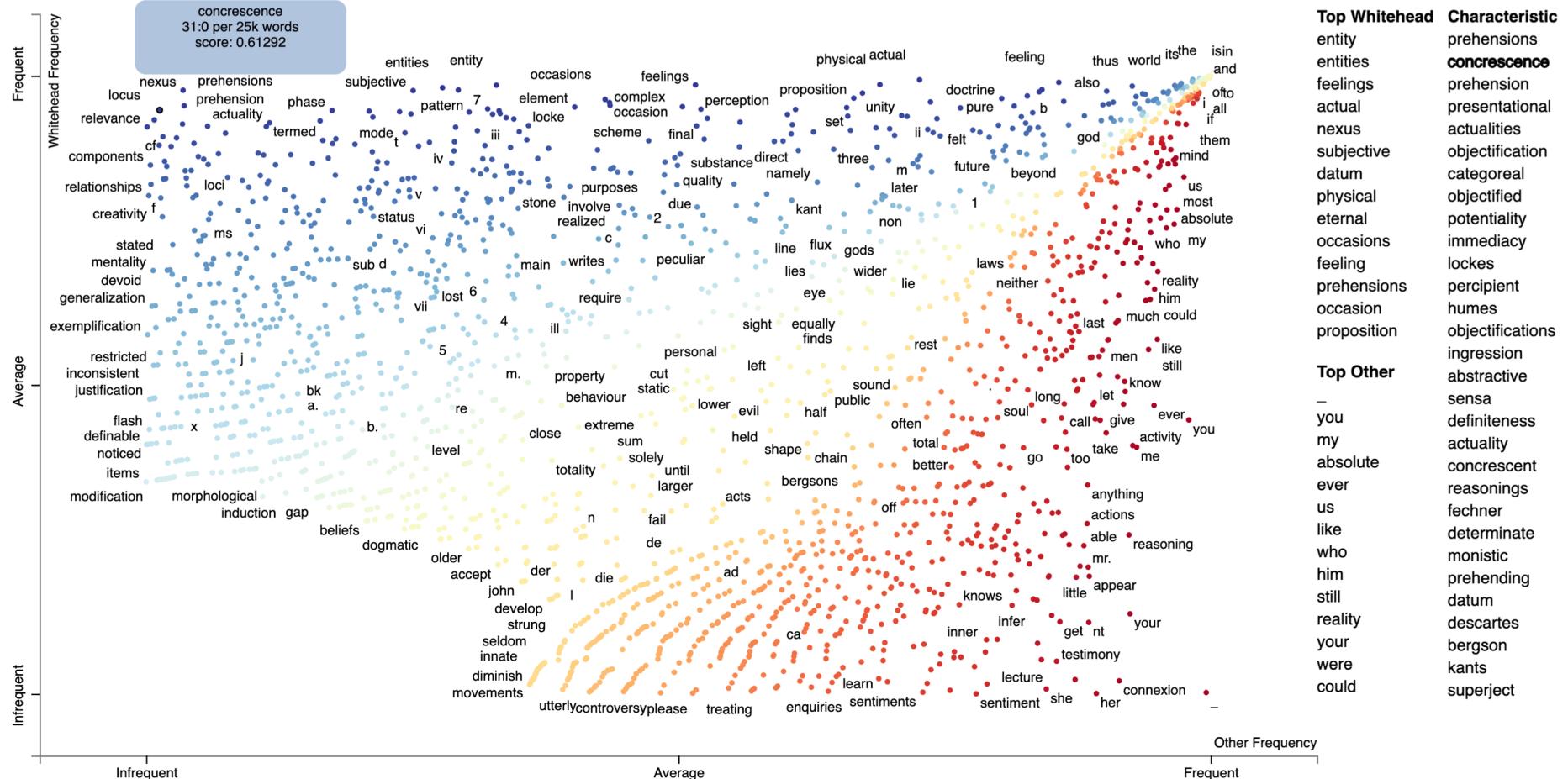


Search the chart

Whitehead document count: 1; word count: 157,800  
Other document count: 2; word count: 126,615

spaCy  
Jason Kessler

## Scattertext of Whitehead vs. Hume and James



## Search the chart

**Whitehead** document count: 1; word count: 157,800  
**Other** document count: 2; word count: 126,615

# spaCy

Search the chart

Whitehead document count: 1; word count: 157,800

Other document count: 2; word count: 126,615

## Term: concrescence

### Whitehead frequency:

31 per 25,000 terms

1000 per 1,000 docs

### Some of the 177 mentions:

### Other frequency:

0 per 25,000 terms

0 per 1,000 docs

Not found in any Other documents.

The coherence, which the system seeks to preserve, is the discovery that the process, or **concrescence**, of any one actual entity involves the other actual entities among its components.

This Category of the Ultimate replaces Aristotle's category of primary substance. Thus the production of novel togetherness is the ultimate notion embodied in the term **concrescence**. These ultimate notions of production of novelty and of concrete togetherness are inexplicable either in terms of higher universals or in terms of the components participating in the con-

22 The Speculative Scheme

crescence.

The analysis of the components abstracts from the **concrescence**.

Thus actual entities are creatures; they are also termed actual occasions/ (ii) That in the becoming of an actual entity, the potential unity of many entities in disjunctive diversity\*—actual and non-actual—acquires the real unity of the one actual entity; so that the actual entity is the real **concrescence** of many potentials.

(iv) That the potentiality for being an element in a real **concrescence**\* of many entities into one actuality!

is the one general metaphysical character attaching to all entities, actual and non-actual; and that every item in its universe is involved in each **concrescence**.

The nexus of actual entities in the universe correlate to a concrescence if it is termed the actual world correlate to that **concrescence**.

(vi) That each entity in the universe of a given **concrescence** can, so far as its own nature is concerned, be implicated in that concrescence in one or other of many modes; but in fact it is implicated only in one mode: that the particular mode of implication is only rendered fully determinate by that concrescence, though it is conditioned by the correlate universe.

This indetermination, rendered determinate in the real **concrescence**, is the meaning of potentiality.

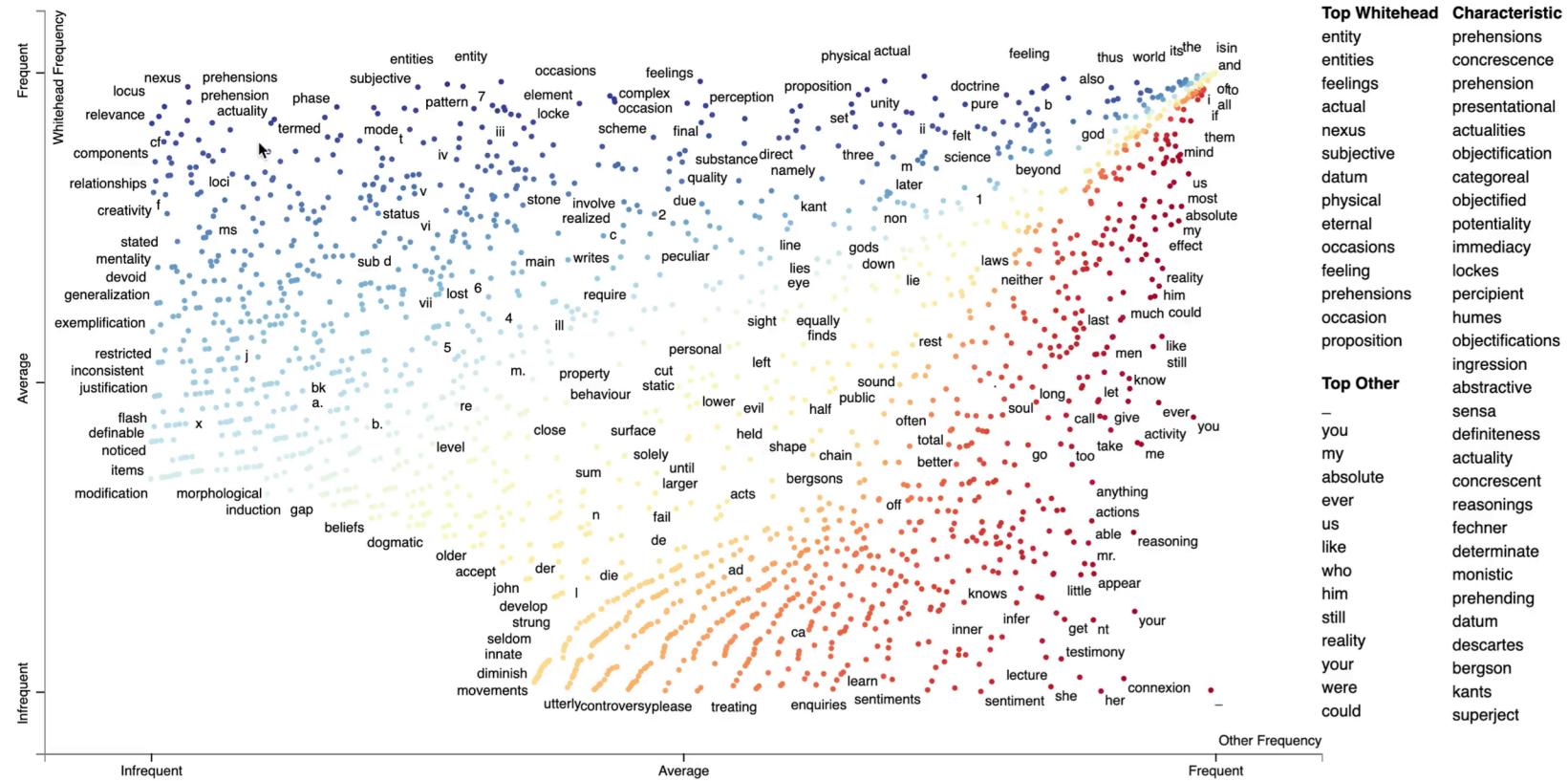
Its being is [35] constituted by its becoming; This is the principle of process/ (x) That the first analysis of an actual entity, into its most concrete elements, discloses it to be a **concrescence** of prehensions, which have originated in its process of becoming.

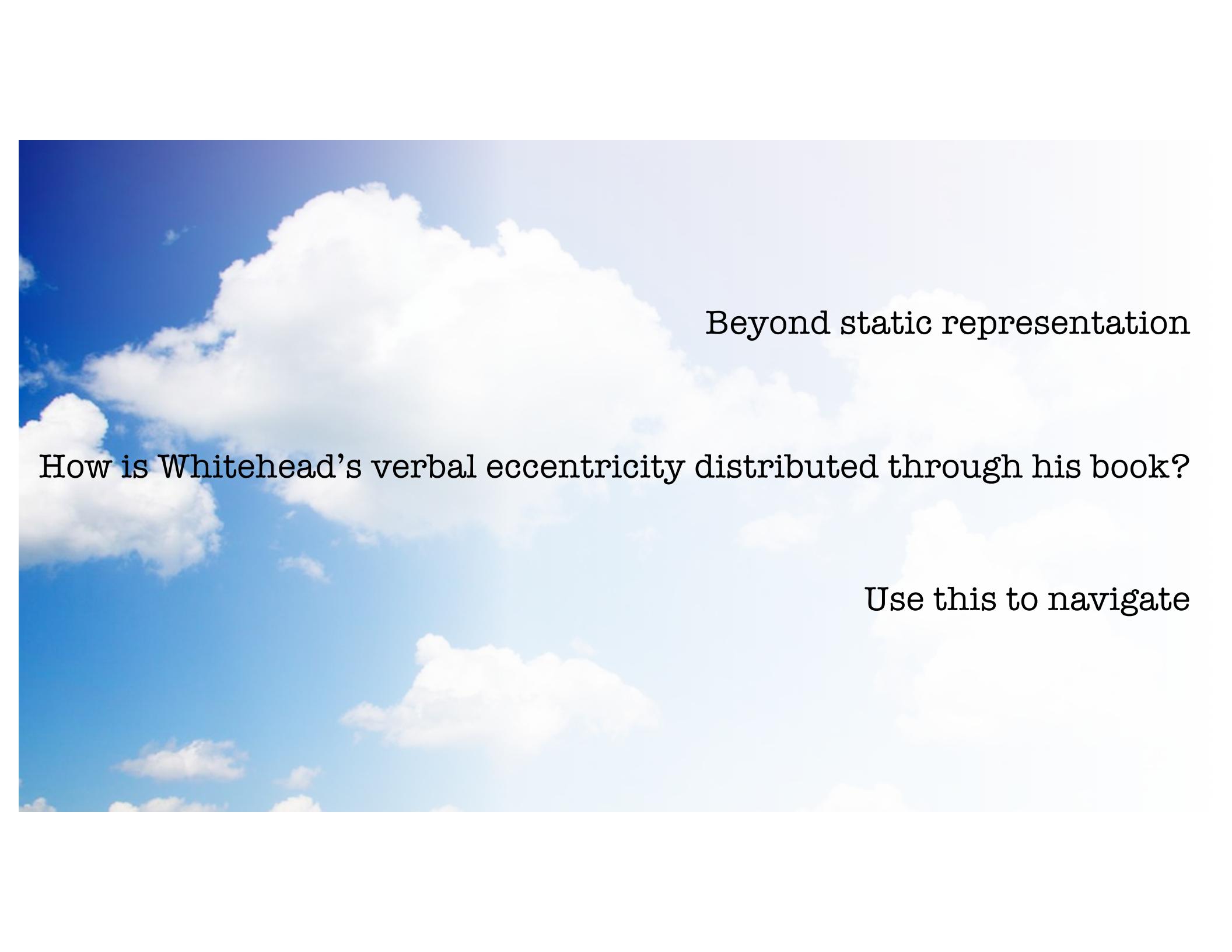
A negative prehension holds its datum as inoperative in the 24 The Speculative Scheme progressive **concrescence** of prehensions constituting the unity of the subject. (xiii) That there are

spaCy

Jason Kessler

Search the chart



The background of the slide is a photograph of a clear blue sky with several large, white, fluffy cumulus clouds scattered across it.

Beyond static representation

How is Whitehead's verbal eccentricity distributed through his book?

Use this to navigate

## **TF-IDF**

Term Frequency \* Inverse Document Frequency

$$\left( \frac{\text{Term Count in document}}{\text{Total Terms in document}} \right) * \log \left( \frac{\frac{\text{Total documents} + 1}{\text{Documents containing the term} + 1}}{\text{Total documents}} \right)$$

## **TF-IDF**

Term Frequency \* Inverse Document Frequency

TF

\*

IDF

TF-IDF often used for document retrieval.

It's been enhanced with hyperparameters for document length.

But document retrieval is not the present use case.

## **TF-IDF**

Term Frequency \* Inverse Document Frequency

$$\text{TF} \quad * \quad \text{IDF}$$

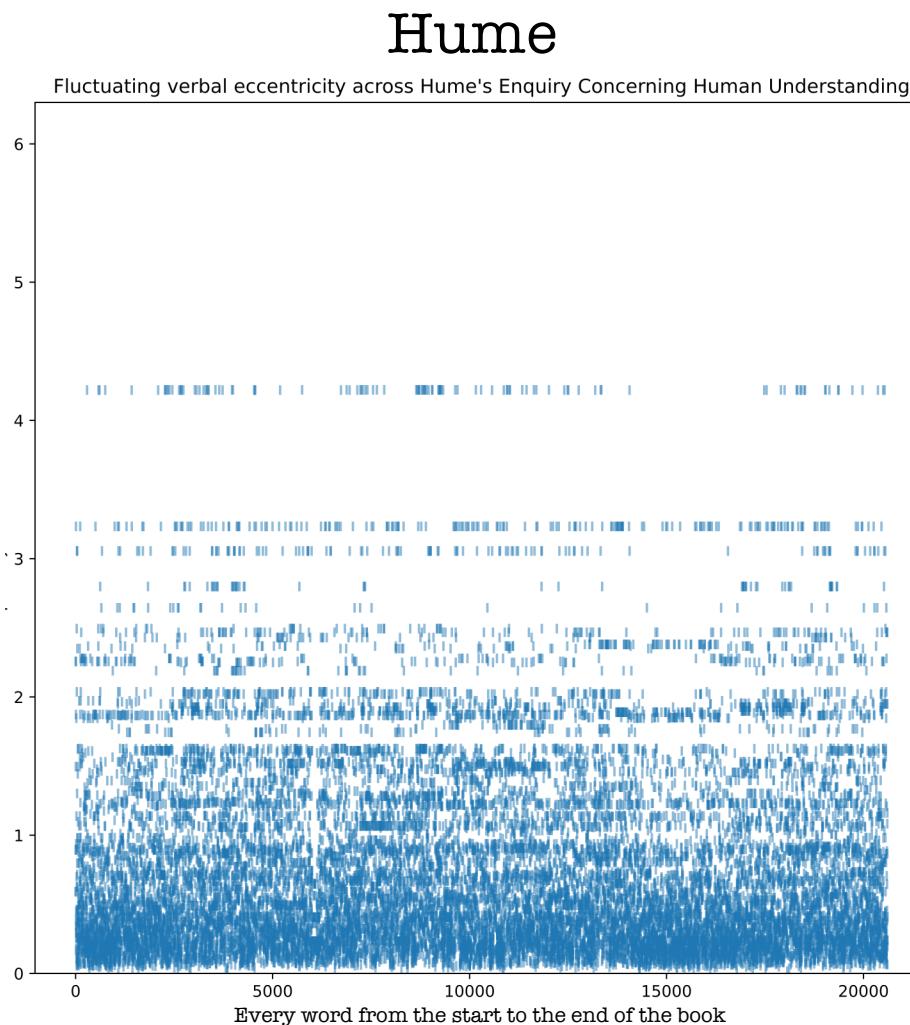
## **Alpha adjusted TF-IDF**

amplifies the effect of rare words when  $\alpha > 1$

$$\text{TF} * \left( \frac{\text{TF}}{\text{TF}^\alpha} \right) * \text{IDF} * \alpha^{(\text{IDF}-1)}$$

Common words are little affected by  $\alpha$

***alpha adjusted***  
***TF-IDF***



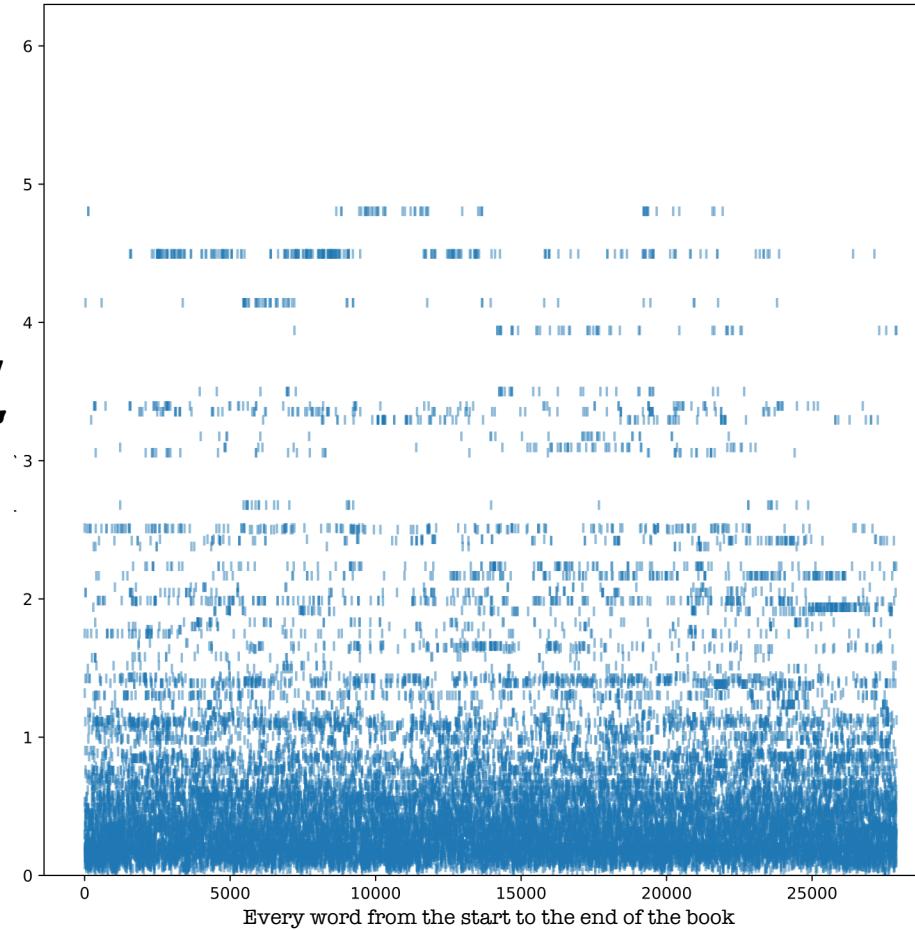
connexion	4.221018
reasonings	4.219499
nature	3.233546
objects	3.055582
qualities	2.798734
enquiry	2.644701
sentiment	2.493891
reasoning	2.467368
senses	2.429436
testimony	2.379780
farther	2.367640
philosophers	2.350014
principles	2.278466
philosophy	2.256766
inference	2.249376
sensible	2.189811
object	2.036775
experience	2.020665
sentiments	1.972153
faculties	1.964636
infer	1.954442
concerning	1.949556
cause	1.926071
effect	1.898431
miracle	1.889445
inferences	1.885650
human	1.873031
mind	1.865688
miracles	1.850808
mankind	1.845740
necessity	1.798002
derived	1.769287
principle	1.744044
regard	1.627200
event	1.626139
imagination	1.625687
instances	1.622959

11K documents from *fetch20newsgroups* corpus used for the outside documents to compute adjusted alpha IDF

# James

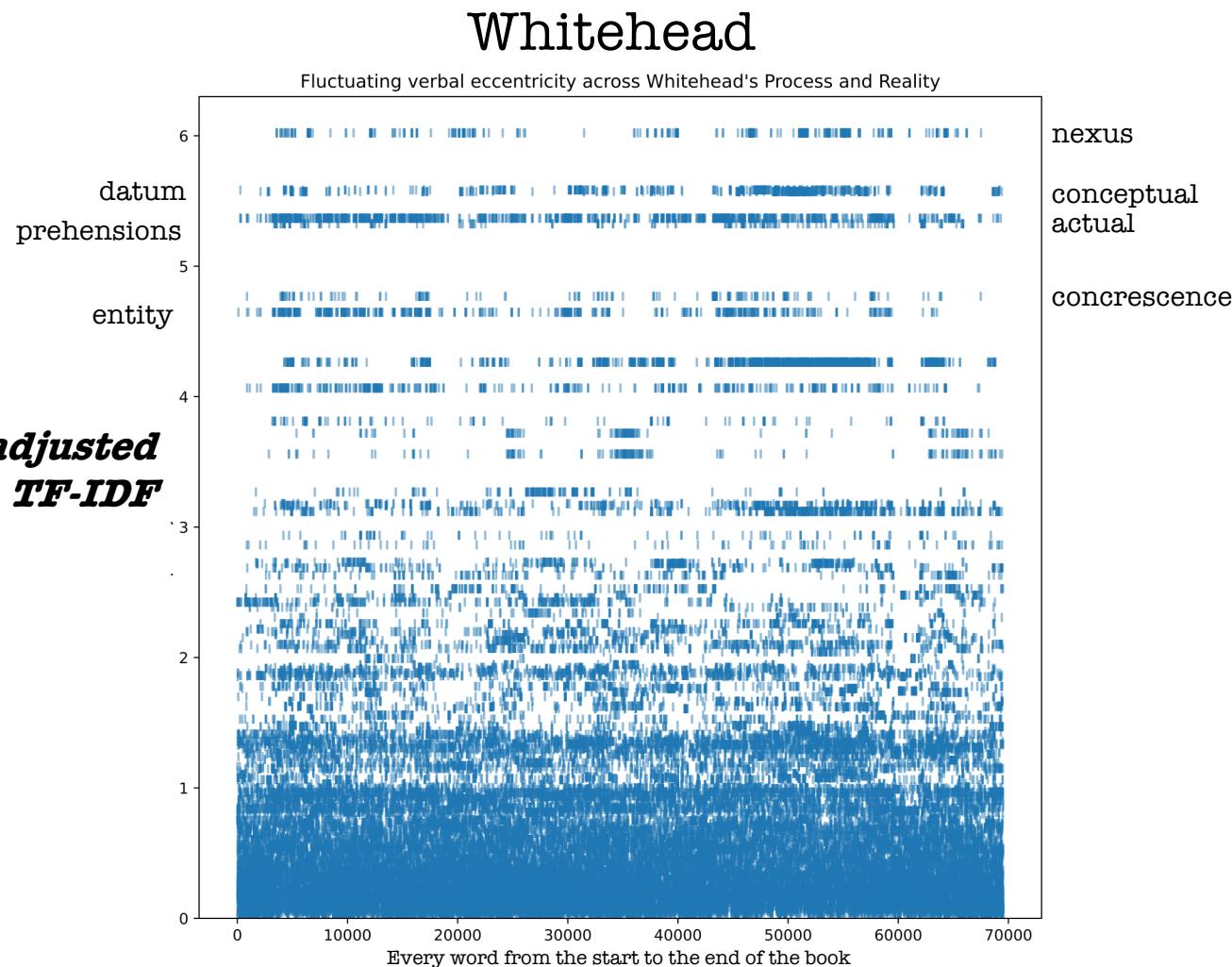
Fluctuating verbal eccentricity across James's Pluralist Universe

***alpha adjusted  
TF-IDF***



11K documents from *fetch20newsgroups* corpus used for the outside documents to compute adjusted alpha IDF

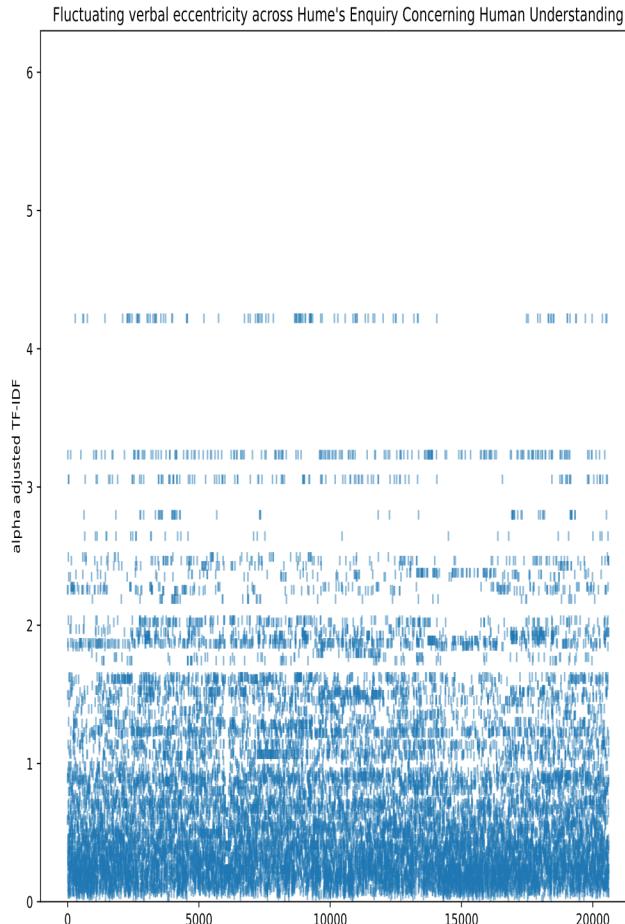
***alpha adjusted  
TF-IDF***



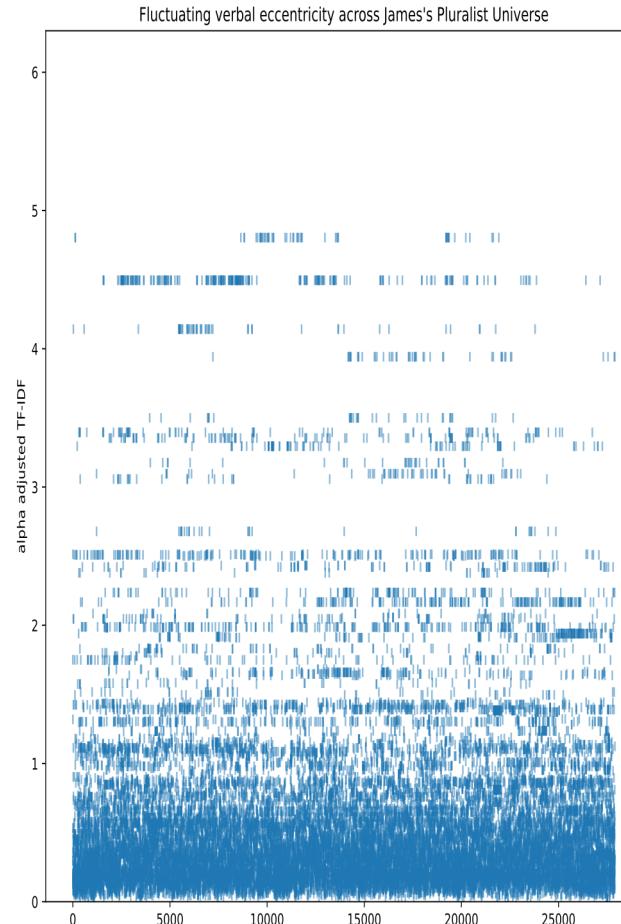
11K documents from *fetch20newsgroups* corpus used for the outside documents to compute adjusted alpha IDF

## Comparing ***alpha adjusted TF-IDF*** of all words instances in three philosophical texts

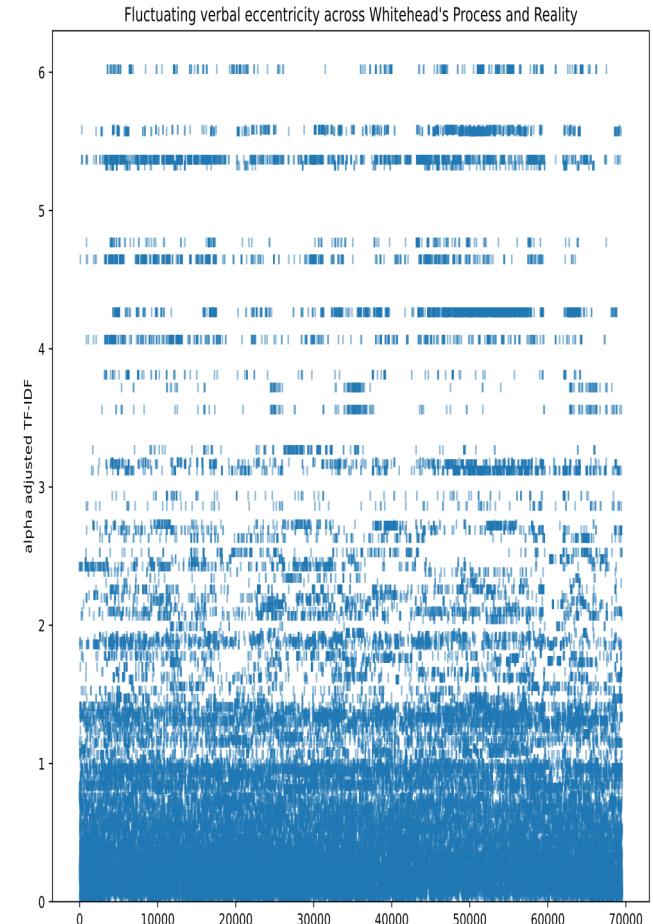
Hume



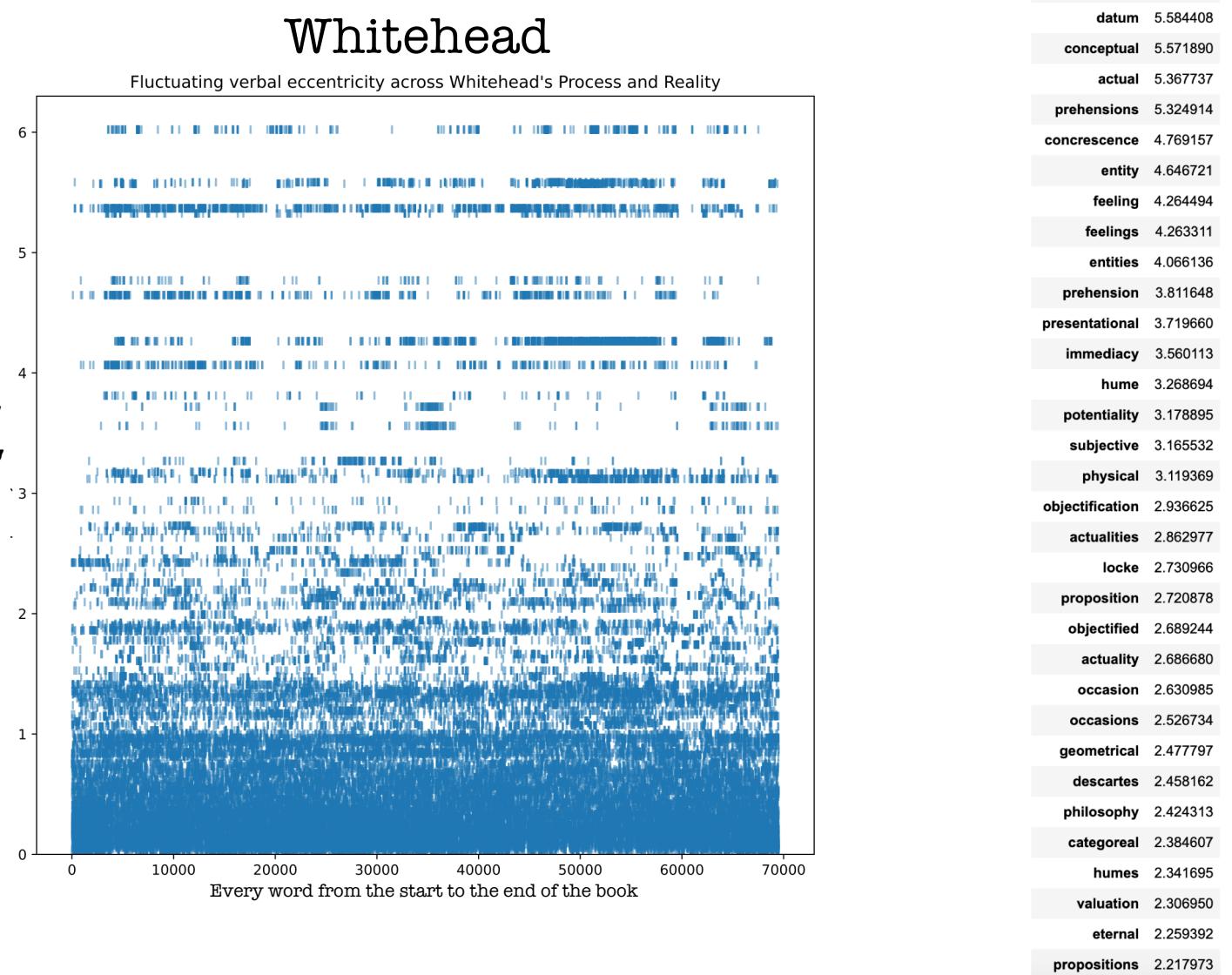
James



Whitehead

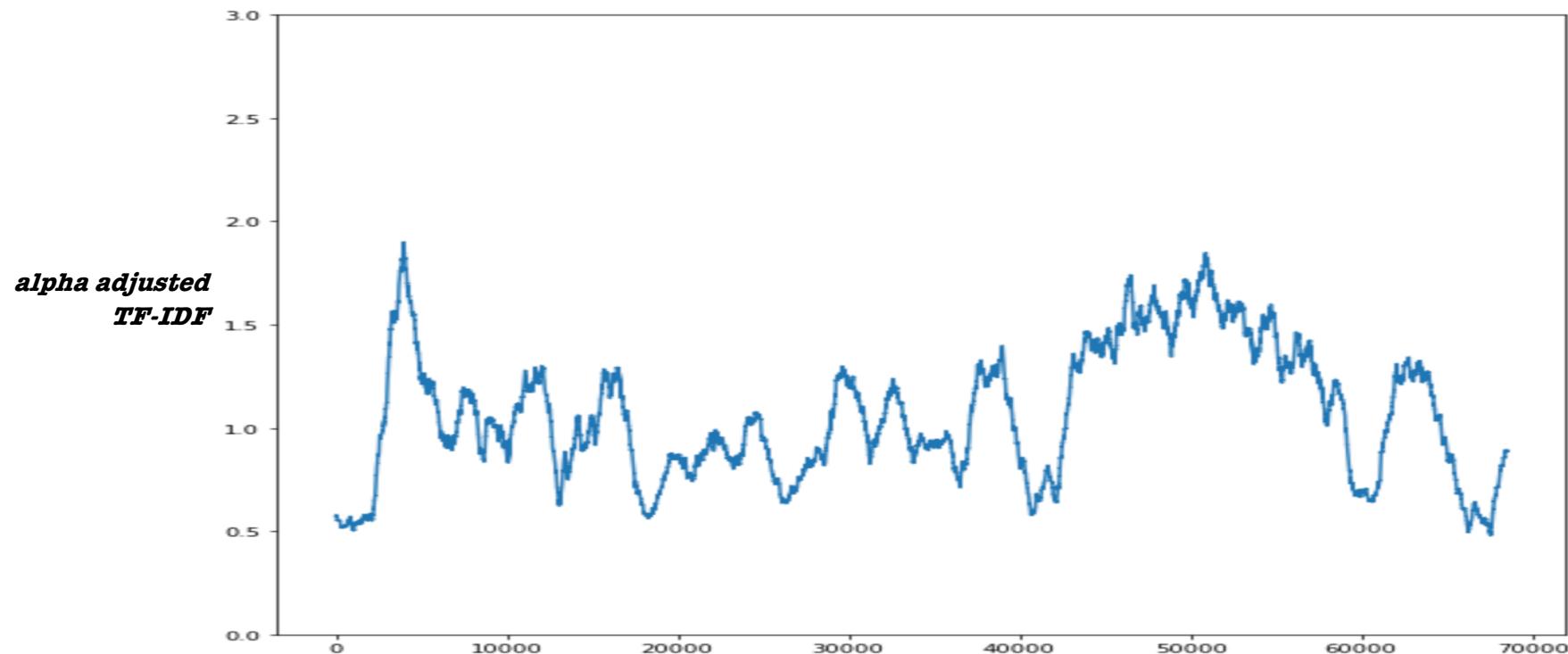


***alpha adjusted***  
***TF-IDF***



# Whitehead

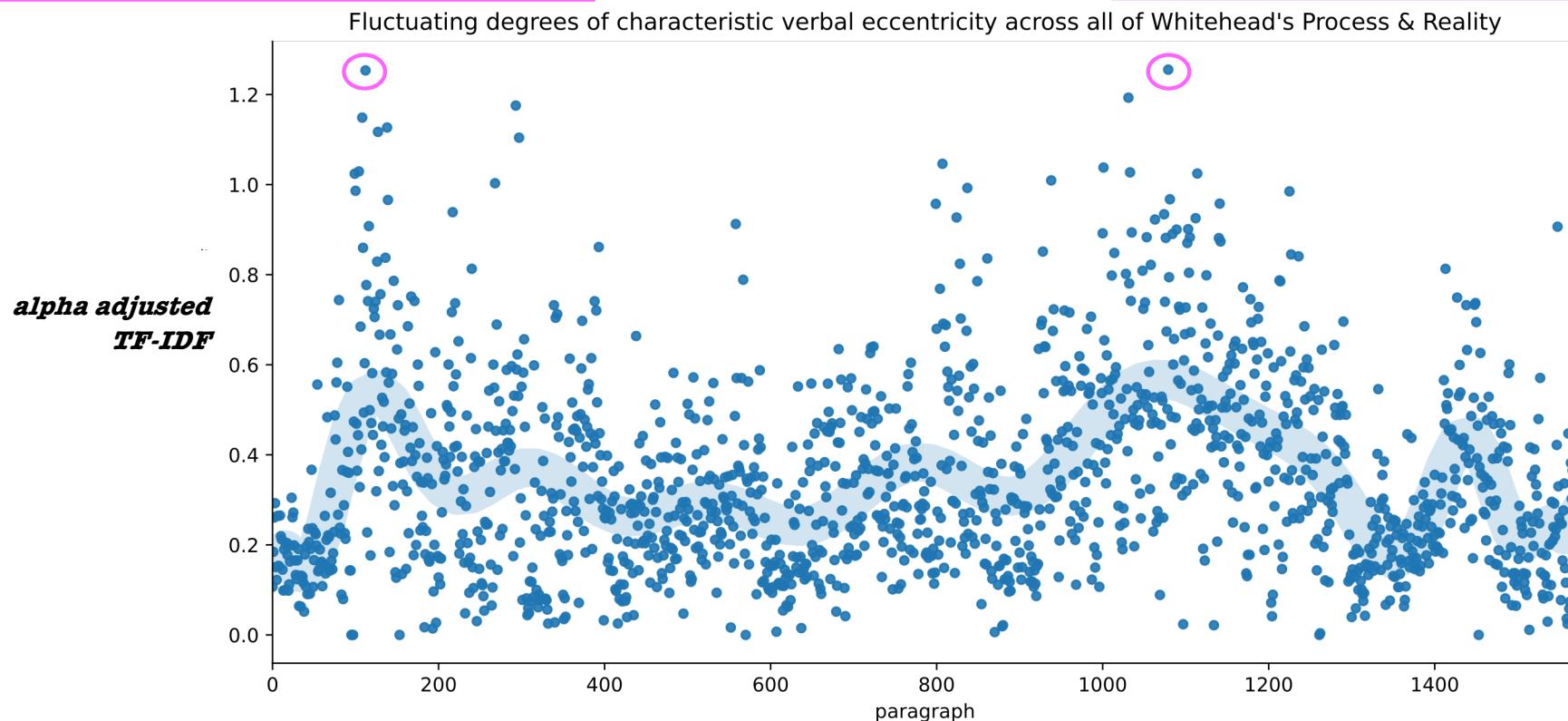
sliding window of 1000 words at a time



Prehensions of actual entities— i.e., prehensions whose data involve actual entities — are termed 'physical prehensions'; and prehensions of eternal objects are termed 'conceptual prehensions/ Consciousness is not necessarily involved in the subjective forms of either type of prehension.

## Whitehead every paragraph

A hybrid physical feeling originates for its subject a conceptual feeling with the same datum as that of the conceptual feeling of the antecedent subject. But the two conceptual feelings in the two subjects respectively may have different subjective forms.

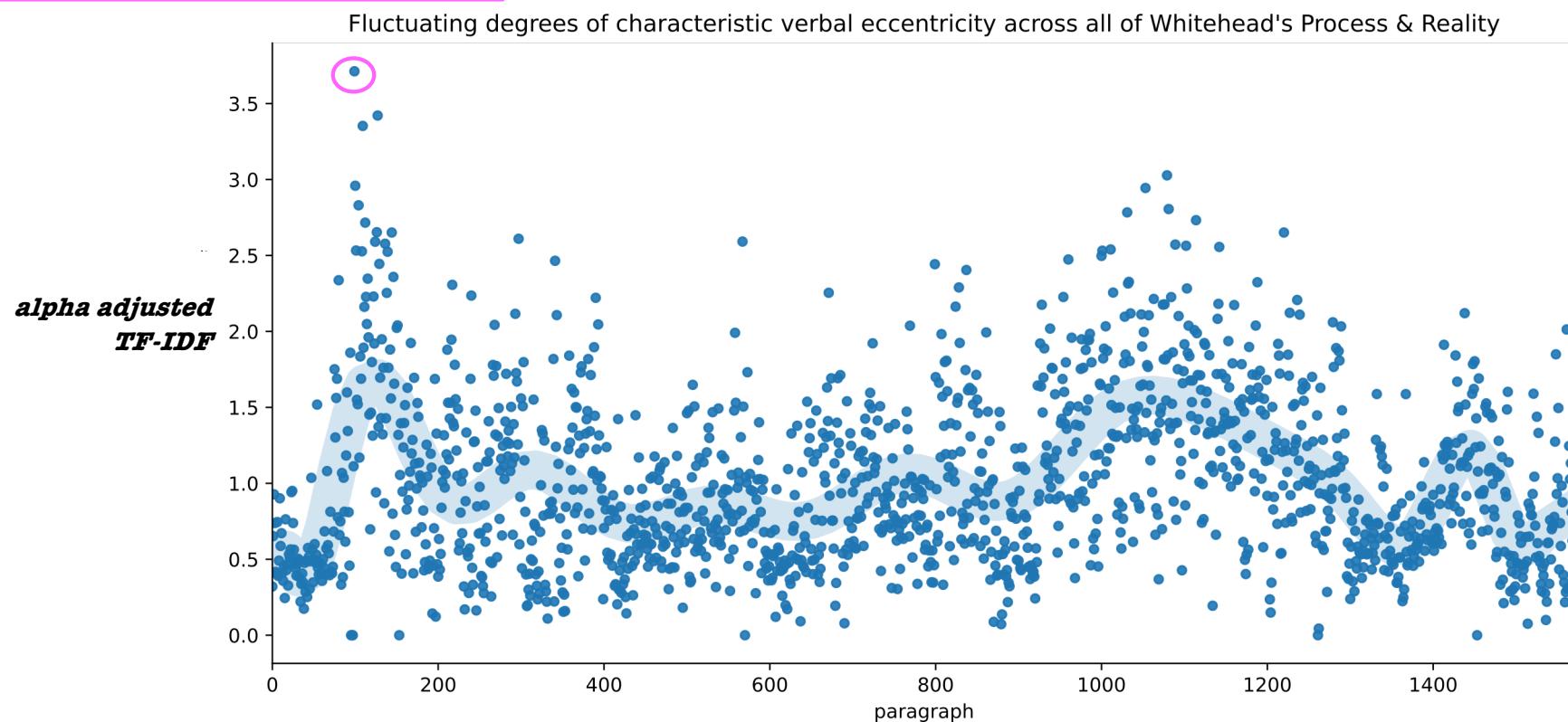


scatterplot with a *lowess* curve

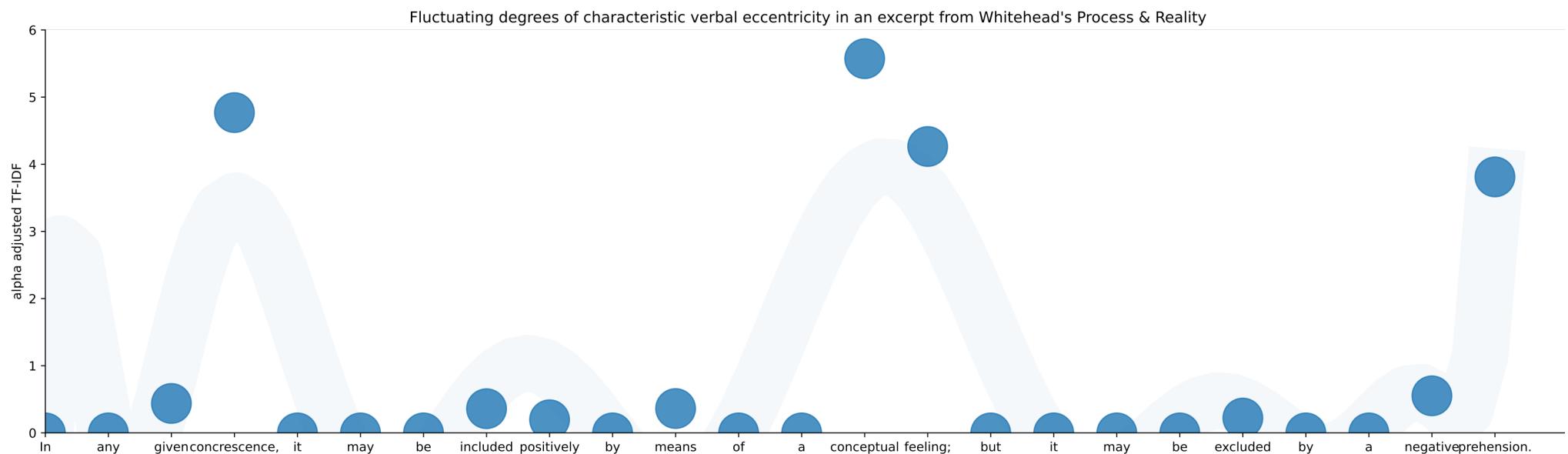
That the actual world is a process, and that the process is the becoming of actual entities. Thus actual entities are creatures; they are also termed 'actual occasions'

## Whitehead

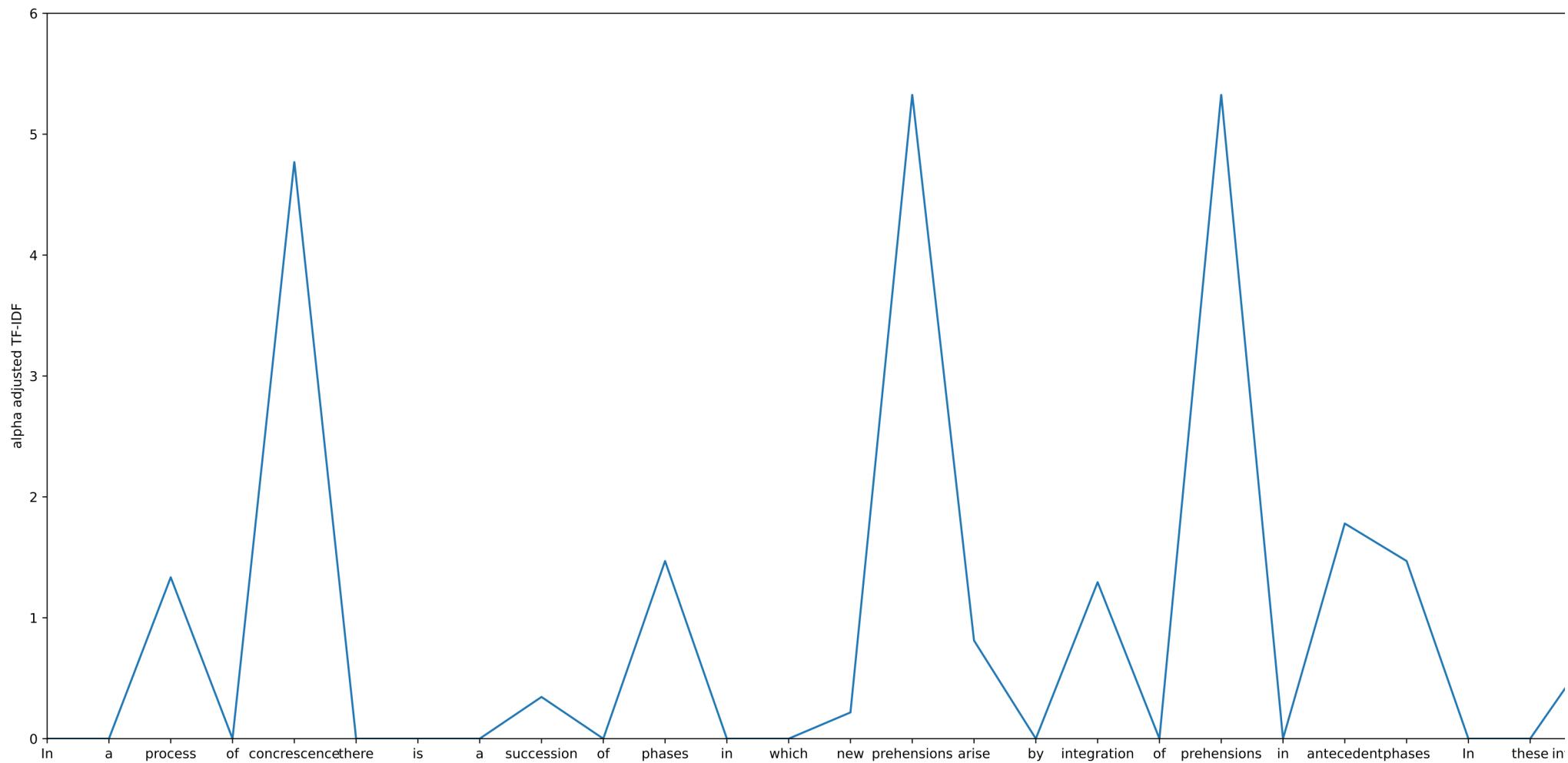
every paragraph  
stopwords excluded



## Possibilities for guiding inflected speech based on flux of characteristic semantic-verbal eccentricity



scatterplot with a high-degree polynomial regression, just as a visual guide



## Applications

Recommender systems:

- Entice book readers and song listeners with book excerpts or song lyric excerpts, based on more specific criteria:
  - Topically consistent passages
  - Self-contained (self-explanatory) passages
  - Verbally unusual, quirky, eccentric, idiosyncratic passages

Automated text-to-speech and algorithmic composition of songs

- Use fluctuating *alpha adjusted TF-IDF* to inflect speech (or melody) based on unusual words prevalent in the document (lyrics)

Education

- New guided approach to navigate difficult, complex, idiosyncratic texts, based on locating (a) topically consistent passages, (b) self-standing (self-explanatory) passages, and (c) idiosyncratic passages.

## Further steps

### Word2Vec

- Train word2vec separately on each book (each philosopher)
  - Select words important to all (mind, concept, ontology) to compare philosophers
    - Use word2vec *wordmover* to see what words are the same vector move away in each philosophers word vector space
    - Compare the same words in each vector space for different ‘positives’ and ‘negatives’
    - Analyze paragraph-to-paragraph topic-flow in terms of word2vec distance between consecutive paragraphs

### CorEx

- Train CorEx separately on each of the book (each philosopher)
  - Select words important to all (mind, concept, ontology) to compare philosophers
    - Use these as *anchor* words and observe the *same* vs. *different* words created in the clustering of words in each book (each philosopher)

More books! More philosophers! More texts of other types!



# New ways to read

‘Language on a holiday’

*Machine learning applied to philosophical texts*

Joshua Banks Mailman

jmailman@alumni.uchicago.edu



Fluctuating verbal eccentricity in an excerpt from Alfred North Whitehead's *Process and Reality*<sup>1</sup>

