What Digital Approaches Can Do for the Study of Intertextuality

Neil Coffee

"Classical Philology Goes Digital" | Feb. 16, 2017 | Potsdam

Intertextuality or Text Reuse?

praeterea iam nec mutari pabula refert
quaesitaeque nocent artes, cessere magistri.
(Vergil *Georgics* 3.548-9)

nec requies erat ulla mali: defessa iacebant
corpora, mussabat tacito medicina timore.
(Lucretius *De Rerum Natura* 6.1178-9)

Hastening the Apocalypse

[There is] "an industry of source-hunting, of allusion-counting, an industry that will soon touch apocalypse anyway when it passes from scholars to computers."

Harold Bloom, 1973,
Anxiety of Influence, 31

Humanistic Goals of Intertextual Study

- Understand authors' art.
- Understand how artistic influence and literary history.
- Understand meaning to original audiences.

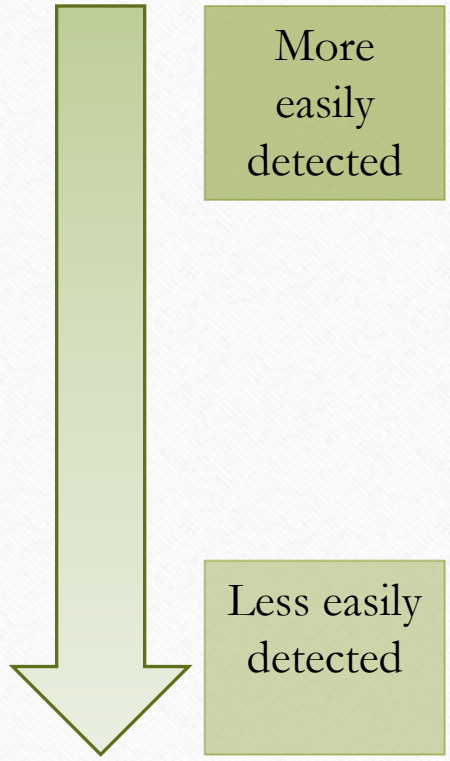
Current Status of Digital Intertextual Research

- Intertextual search engines work, though they need to be improved
- Digital notation of intertexts becoming a possibility
- Where to go from here?

Defining and Finding Intertextuality

Language features by detectability

Language Feature
Exact word
Lemma
Markers of quotation or allusion
Meter
Syntax
Sound
Section boundaries / narrative structure
Semantic relatedness (synonym, antonym, metonym)
Context



Definition vs. Description

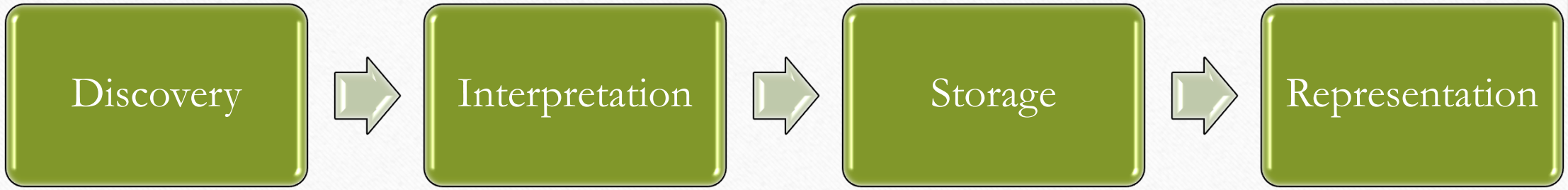
	Features	Count	% of total
1.	Two-word (exact word or lemma) identity	146	67%
2.	One identical word + semantic context	27	12%
3.	One identical word + synonym	16	7%
4.	Semantic context only	13	6%
5.	Synonyms only	12	6%
6.	One identical word + syntax	2	1%
7.	One identical word + sound	2	1%
	TOTAL	218	100%

Digital Tools for Intertextual Search

- *Perseus*, *PHI*: string searches, Boolean operators
- *Filum*: edit distance, Latin prose and poetry
- *Musisque Deoque*: lemma, meter, word order; Latin poetry
- *Tesserae*: lemma, word frequency, meaning, sound; Greek and Latin prose and poetry
- *TRACER*: search by numerous features; downloadable software package

How We (Can) Do Intertextual Research

Intertextual Reading / Research Process



Research Scenario 1 – Traditional Unassisted Reading

- Discovery and Interpretation
  - Recall of passage while reading. Reflection on parallels and interpretation.
- Storage
  - Article, commentary, monograph. Some loss of information due to publication formats.
- Representation
  - Article, commentary, monograph.

Research Scenario 2 – Reading and Targeted Digital Search

- Discovery
  - Targeted search for passage found while reading.
- Interpretation
  - Reflection on parallels and interpretation.
- Storage
  - Article, commentary, monograph. Some loss of information due to publication formats.
- Representation
  - Article, commentary, monograph.

Research Scenario 3 – Global Automatic Search

- INTERPRETATION
  - Selection of research query (e.g., which texts to compare)
  - Search algorithm for matching selected, representing (minimal) interpretive schema.
- DISCOVERY
  - Unsupervised detection of parallels by algorithm.
- INTERPRETATION
  - Reflection on parallels and interpretation.

Research Scenario 3 – Global Automatic Search

- STORAGE
  - Search results can be stored online for public access. Curated results stored in article, commentary, monograph. Some loss of information due to publication formats.
- REPRESENTATION
  - Article, commentary, monograph.

Agenda for Tool Development

1. Develop a universal standard for denoting intertexts.
2. Implement standard in textual editing and annotation environments.
3. Implement standard on search engines for search and delivered results.
4. Enable collation of marked published intertexts from editing environments.
5. Develop improved search using marked intertexts as benchmark.

Develop a universal standard for denoting intertexts

- Now supplied by Berti, M., C. Blackwell, et al. 2016. "Documenting Homeric Text-Reuse in the Deipnosophistae of Athenaeus of Naucratis." *BICS* 59(2): 121-139.
- 6-part standard, referring to source locus (e.g. *Iliad*), target locus (e.g., *Aeneid*), specific language used. Uses CTS citation standard to refer to loci. Includes unique identifier *for the intertext* (CITE-standard).
- Possible augmentation:
  - intertexts referring to more than two loci (as in roses and lilies example)
  - Incorporation of *type* of intertext. E.g., SAWS ontology for types of textual relations
    - (e.g., "isLooseRenderingOf", and "isCloseRenderingOf")

Implement standard in textual editing and annotation environments

- Presumably under way at Perseids. Similar integration for DLL?
- Intertexts can be marked up in copies of text that can be published and compared.
- Form of micro-publication.
- Creation of a clearing-house of intertexts so that all can be easily found and none are lost or repeated.
- Include where possible collations of existing published intertexts in commentaries and other sources
- Enable search across published intertexts from within editing platform.

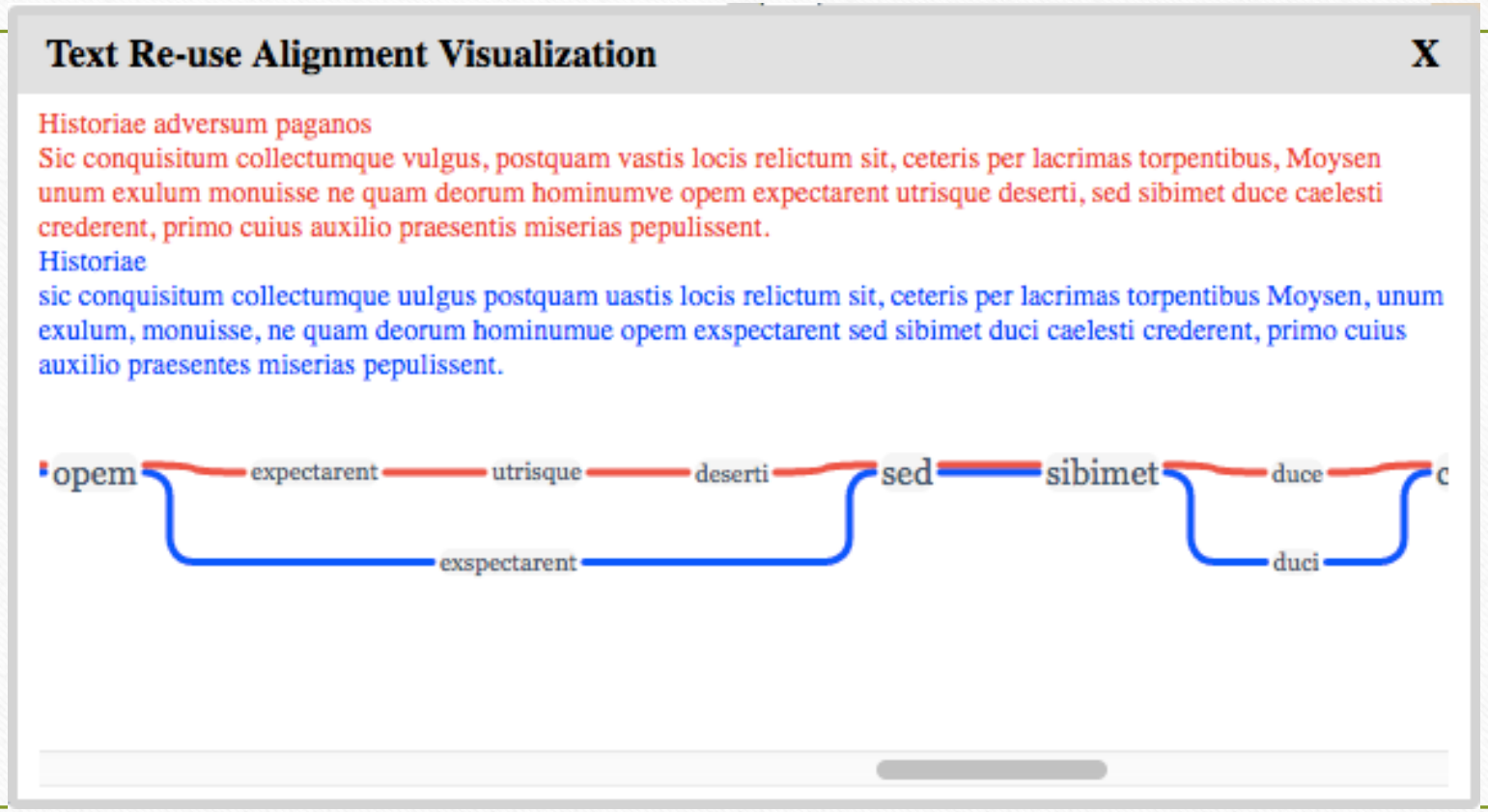
Implement standard on search engines for search and delivered results

- Tesserae and Musisque Deoque have a project this year to implement this standard. Include API on all search engines.
- Would allow for searches across multiple platforms, comparison and collation of results.
- APIs could be used by Perseids and DLL to incorporate intertextual search and markup into their editing and annotation platforms.

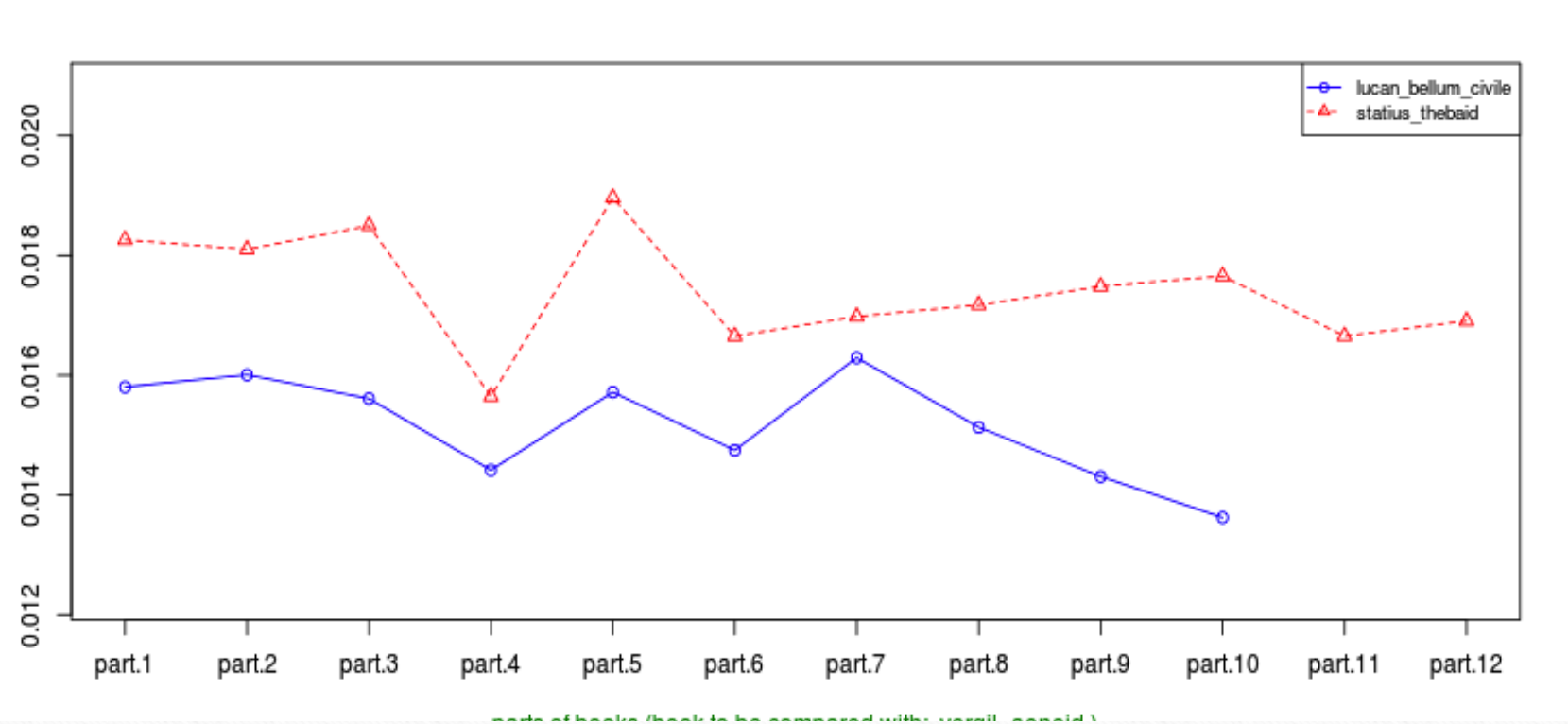
Enable collation of marked published intertexts from editing environments

- These could be used for literary research.
- Used for study of intertextuality, as a gold standard against which to test search engines.
- Used for visualizations, such as those provided by TRACER.

TRACER visualization
Orosius and Tacitus



Tesserae experimental visualization
Relative use of *Aeneid* by Statius (red) and Lucan (blue) by book



Develop combined feature search using marked intertexts as benchmark

- Ideal search will:
  - Be fully customizable by features
  - Offer best approximation settings for finding intertextuality by different features.
- Using existing intertexts entered as annotations can be used as benchmark sets to test different feature combinations, adjust for language, genre, etc.

Research Scenario 4 – Search, Repository, Visualization

- INTERPRETATION
  - Selection of research query (e.g., which texts to compare)
  - Search algorithm for matching selected, representing (minimal) interpretive schema.
- DISCOVERY
  - Unsupervised detection of parallels by algorithm.
- INTERPRETATION
  - Reflection on parallels and interpretation.

Research Scenario 4 – Search, Repository, Visualization

- STORAGE
  - Entirety of research results can be stored online for public access in different formats, with no loss of information.
- REPRESENTATION
  - Conclusions can be represented in a variety of digital formats, separately or aggregated with other research. This includes visualizations. Can also be presented in traditional formats of article, commentary, monograph.

Thanks

- University of Potsdam, Karen Blaschka, Monica Berti
- Gregory Crane, Lavinia Galli-Milic, Damien Nelis, Matteo Romanello
- Tesserae team. tessserae.caset.buffalo.edu
- University of Venice / *Musisque Deoque*
- NEH ODH, Swiss National Science Foundation, Fondation Maison des Sciences de l'homme, University at Buffalo, Notre Dame University