eTRAP [electronic Text Reuse Acquisition Project]

A research group implementing the A.C.I.D. eHumanities paradigm

EMILY FRANZINI, HUMANIST
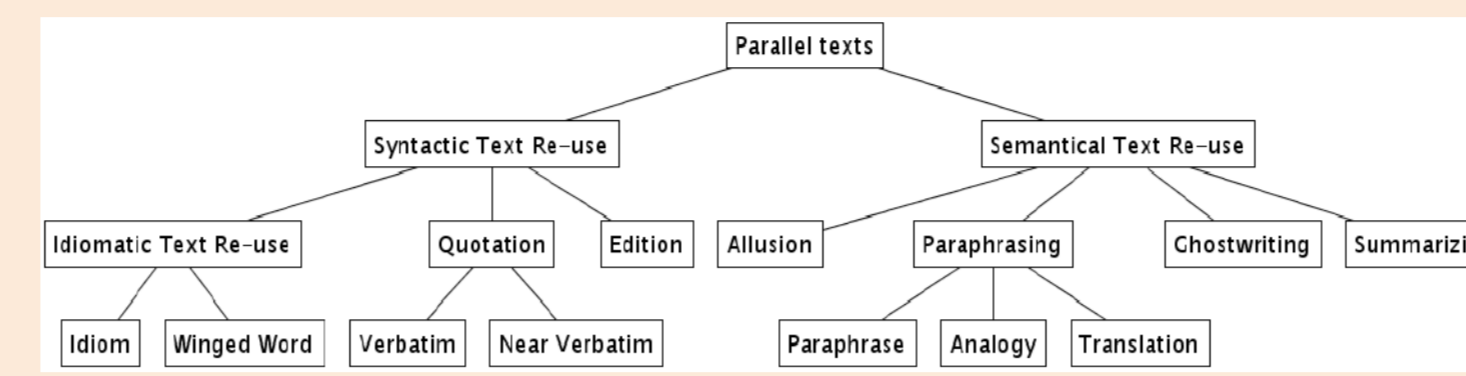
Biblia
τὰ βιβλία τὰ ἅγια
The Bible
La Bibbia
die Bibel

Research Questions

The re-use of text across languages is a gold mine for **transcultural** studies. When observing translated text,…
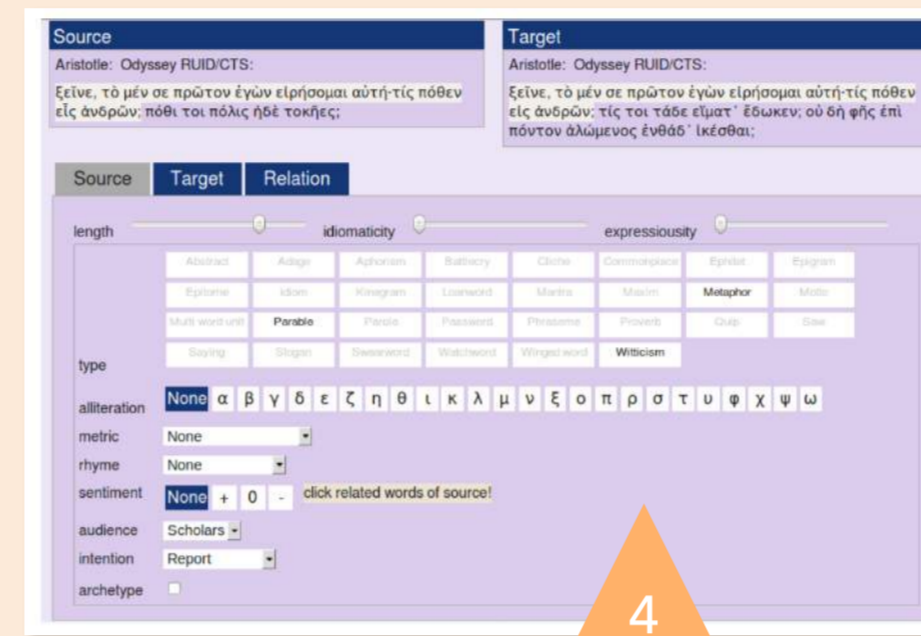
▸ …what is the scale of divergence from one text to the other?

▸ …what is the divergence caused by? Translator competence or linguistic, cultural, ideological and political norms?

▸ Could machine translators be used to rid a translation of any external contexual influences?

D(iversity)

The many styles of text re-use.

Proposed methodology: Text re-use browser prototype.

MARCO BÜCHLER, COMP. SCIENTIST

Research Questions

▸ Which are the determinable **minutiae** of text re-use and which of them can be calculated by machines?

▸ What can't, by contrast, be calculated?

▸ How can text re-use analysis be improved by the calculation of paradigmatic relations?

I(nteroperability)

Annotation; storage; encoding standards. How to **homogenise** datasets in order to ensure interoperability and **minimise** data preparation/conversion?

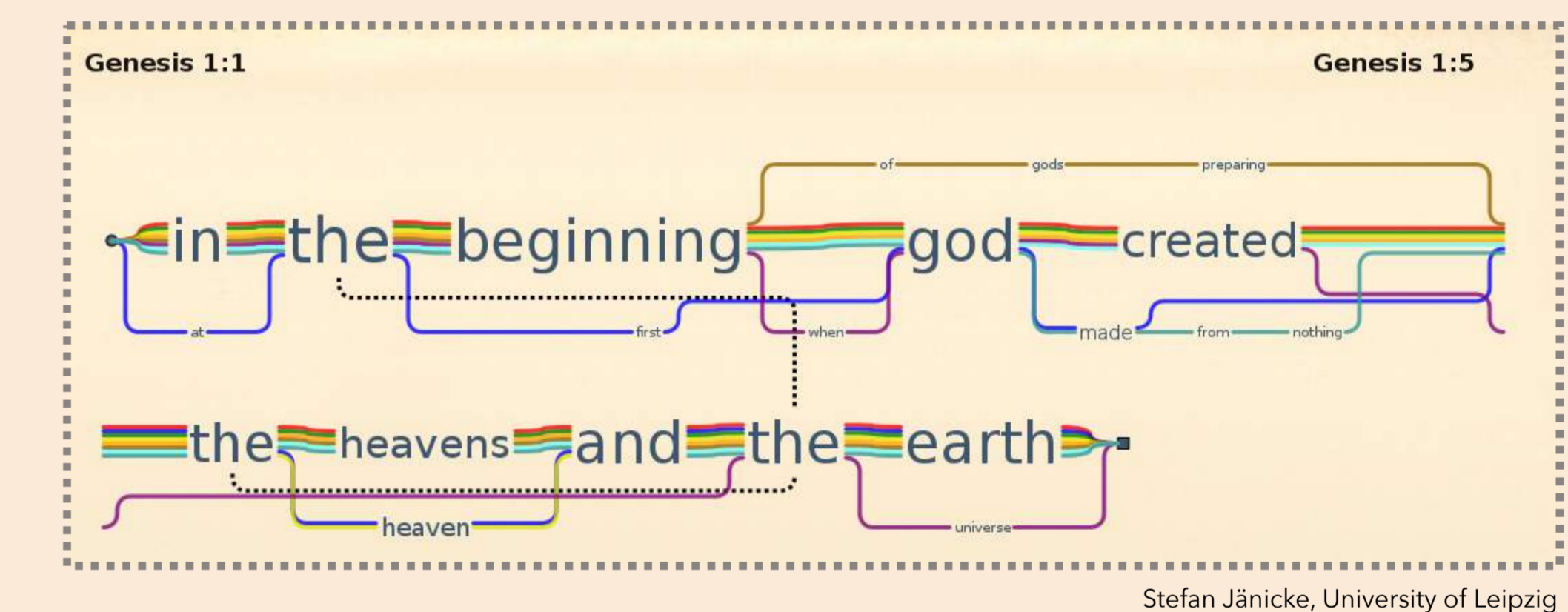The example below shows three ways of annotating text re-use.

Proposed methodology: store text data in Canonical Text Services and links in RDF.

"IN THE BEGINNING GOD CREATED THE HEAVENS AND THE EARTH"
GENESIS, BOOK 1

A(cceptance)

Information overload vs. information poverty. Packaging overload while retaining detail.

Proposed methodology: text graph variation visualisation (TRAViz) and URN notation (CTS/CITE Architecture).

Stefan Jänicke, University of Leipzig

Research Questions

When observing a sentence or word that has been used in multiple texts, one may ask:

▸ How to **organise** heterogenous and **complex data**?

▸ What are the differences and similarities, and what do they tell us about the texts?

▸ How does one quantify and classify the similarities and the differences?

C(omplexity)

700,000 manually collectect biblical references (*Biblindex*);

150 million words (*archive.org*);

40 million Ancient Greek and Latin words (*Open Philology Project*);

52 editions of English Bible translations (*mysword.info*).

Proposed methodology: Historical text re-use detection community building.

Research Questions

Can the study of text re-use…

▸ …on high definition **images** of manuscripts enhance comparable research on **digital text**?

▸ …tell us something about the accuracy and confidence of variation? Which authors quote the source more literally and why?

▸ … show us that the variation noise tells us something about the source of the variation? What is the nature of the transmission contamination?

MARIA MORITZ, COMP. SCIENTIST

GRETA FRANZINI, HUMANIST

http://etrap.gcdh.de