

# eTRAP [electronic Text Reuse Acquisition Project]

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

EMILY FRANZINI, HUMANIST

BIBLIA  
τὰ βιβλία τὰ ἄγία  
THE BIBLE  
LA BIBBIA  
DIE BIBEL

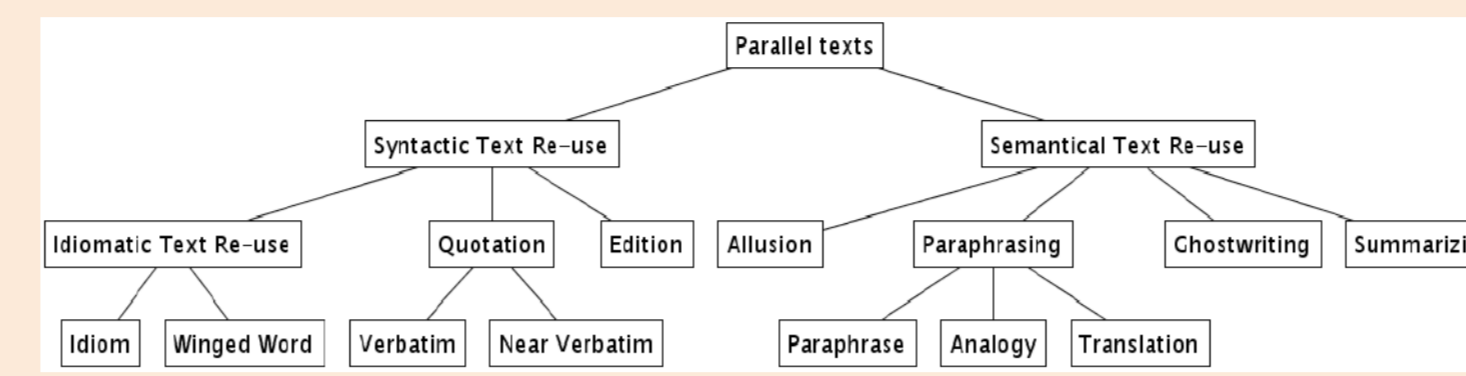
## Research Questions

The re-use of text across languages is a gold mine for **transcultural** studies. When observing translated text,...

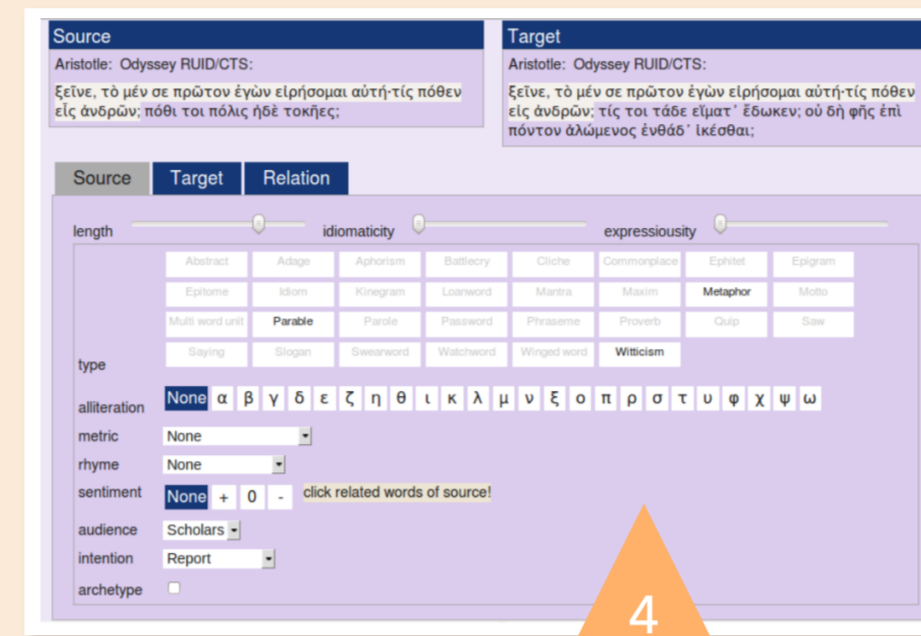
- ▶ ...what is the scale of divergence from one text to the other?
- ▶ ...what is the divergence caused by? Translator competence or linguistic, cultural, ideological and political norms?
- ▶ Could machine translators be used to rid a translation of any external contextual influences?

## 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.

Cit-quote-bibl	blockquote	bibl without quote
<pre>&lt;qt&gt;   &lt;quote&gt;du/o ku/nas a)rgoi\ ei   (/ponto&lt;/quote&gt;   &lt;bibl n="Hom. Od. 2.11"&gt;Od. 2.11&lt;/bibl&gt; &lt;/qt&gt;</pre>	<pre>&lt;quote rend="blockquote"&gt;   - &lt;line&gt;   a)rxou= d' i(stame/nh e)/pea   ptero/anta proschu/da   &lt;bibl n="Hom. Il. 4.92"&gt;Il.   4.92&lt;/bibl&gt;   &lt;/line&gt;   - &lt;line&gt;   a)ll' a)/ge nu=n ma/stiga kai\ h   (ni/a sigalo/enta   &lt;bibl n="Hom. Il. 5.226"&gt;Il.   5.226&lt;/bibl&gt;   &lt;/line&gt;   &lt;/quote&gt;</pre>	<pre>&lt;p&gt;   [...]a)nti\ tou= proe/pinon. kuri/ws   ga/r)sti tou=to propi/nein, to\ e   (te/rw) pro\ e(autou= dou=nai piei=n.   kai\ o( *jodusseu\s de\ para\ tw=  *   (omh/rw)   &lt;bibl n="Hom. Od. 13.57"&gt;Od.   13.57&lt;/bibl&gt;   [...]</pre>

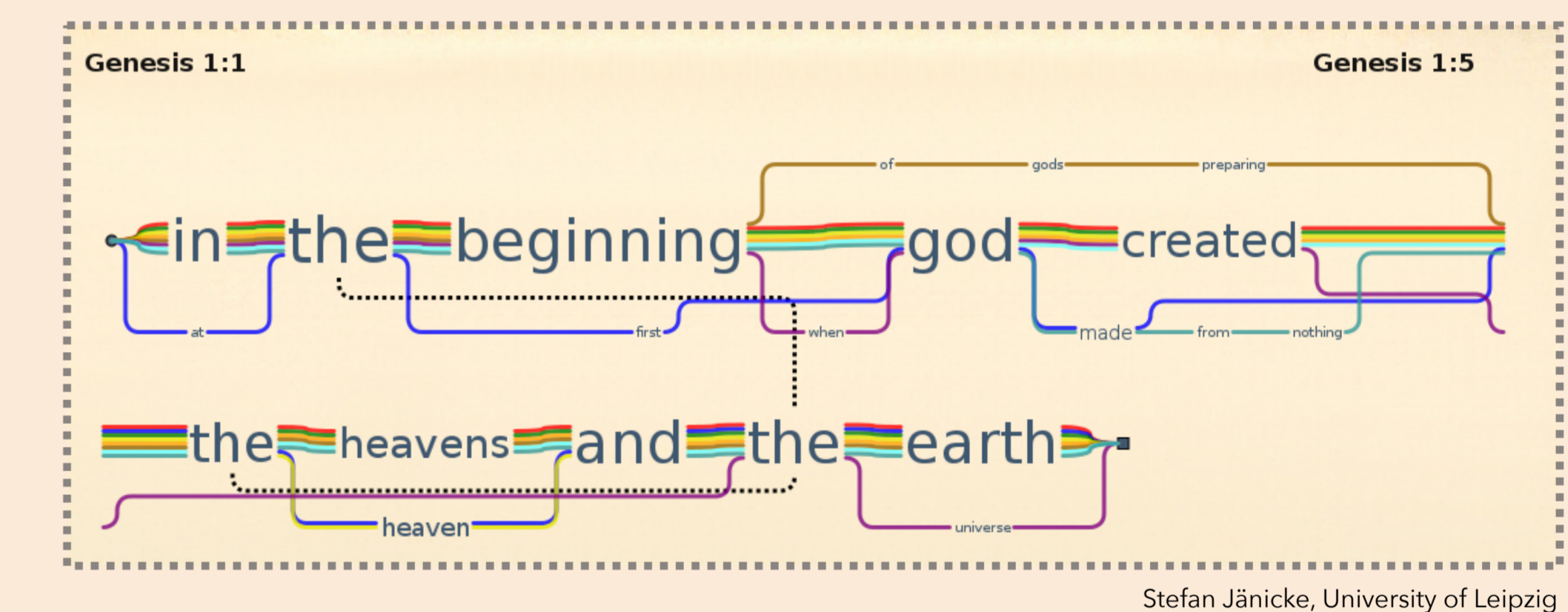
**Proposed methodology:** store text data in Canonical Text Services and links in RDF.

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

## 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).



## C(omplexity)

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

**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