

Citation Practices in Jerome's Letters as *vestigia* of Late Antique Identity Construction

Marie Revellio

Universität Konstanz, Department of Literature/ Latinistik; Digital Humanities Cooperation
Technische Universität Darmstadt & Universität Konstanz

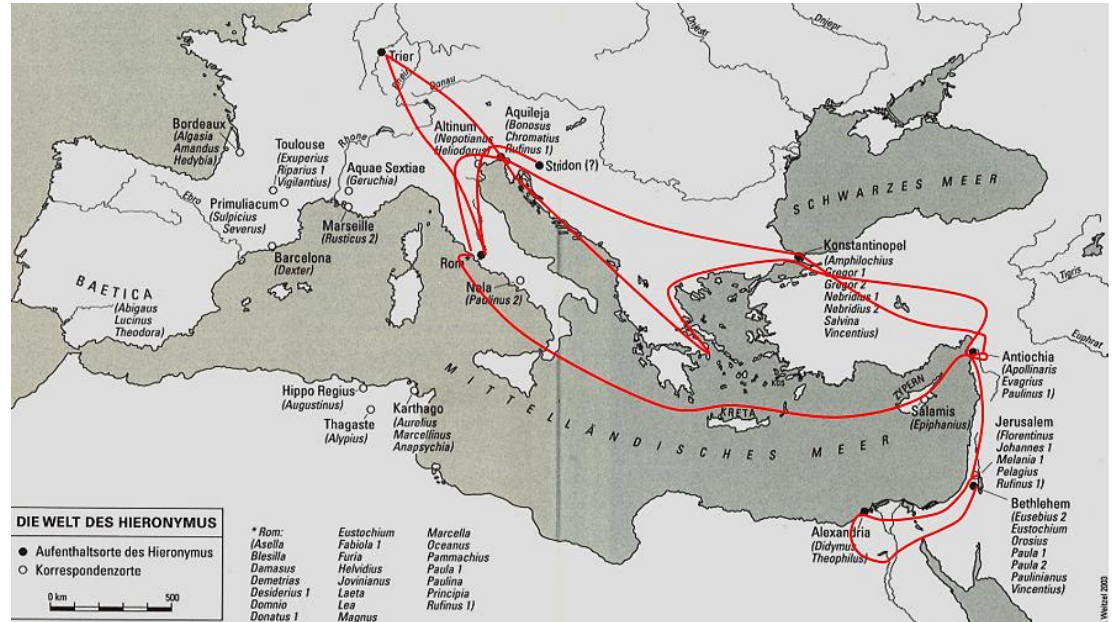
Outline

1. Jerome – Background and motivation
2. Research question & state of research
3. Data
4. Research design
 - 4.1. Correct classification
 - 4.2. Evaluation of classification system
 - 4.3. Optimization rules
5. First steps of analysis (*Tesserae*)
 - 5.1. Filter criteria
 - 5.2. Basic code
6. First results & further plans
7. Discussion

Sophronius Eusebius Hieronymus (c. 347 – 420) – Mediator between the eastern and western Roman Empire



Dürer, Der heilige Hieronymus im Gehäuse (1514).



Fürst, Hieronymus (2016); Jerome`s travel routes (in red) are mine.

Œuvre:

commentaries and biblical studies, homilies and tractates, letters, polemical writings, translations, novelistic lives of the saints, *De viris illustribus*.

Research motivation – Jerome's strong focus on literature

Jerome's dream – a paradigm of the struggle with classical education in late antiquity:

cum subito raptus in spiritu ad tribunal iudicis pertrahor, ubi tantum luminis et tantum erat ex circumstantium claritate fulgoris, ut proiectus in terram sursum aspicere non auderem. interrogatus condicionem Christianum me esse respondi. et ille, qui residebat: 'mentiris', ait, 'Ciceronianus es, non Christianus'. (...) ego, qui tanto constrictus articulo uellem etiam maiora promittere, deiurare coepi et nomen eius obtestans dicere: 'domine, si umquam habuero codices saeculares, si legero, te negauit'.

Jer. ep. 22,30,3-5 (CSEL 54, 190-1)

- (religious) identity is determined by the reading of literature

Research verdict:

„Jerome the literary exhibitionist“ (Cain 2013a:20)

„Jerome quotes poetical lines to an extent unprecedented in other writers, pagan as well as Christian, with the exception of Cicero in his philosophical works, Lactantius and Augustine“ (Hagendahl 1958:298)

The project

Background:

The project regards intertextual relations as

- a literary stage to deal with questions concerning cultural hybridity
- markers for the literary processing of a manifold cultural transformation

The aim is to create a typology of the citations and to work out the author`s citation technique.

Main research questions:

How does Jerome weave citations and allusions into his texts?

Are there differences in marking/ pondering/ valuating the source texts in respect of their cultural/ religious background? Does this point out for a narrative strategy?

To which extent is the phenomenon of intertextuality

- a literary instrument to construct late antique identity between Christian and classical-pagan culture/ literature/ beliefs?
- entangled to Jerome`s conception of Christian authorship?

Research object – Vergil`s poetry in Jerome`s letters

- Jerome is at his best in his letters “where he gives free rein (...) to his rhetorical skill” (Hagendahl 1958:314)
- Jerome does mark his reused passages in the letters less than in other of his writings
- citations from poetry consist of relatively distinct language
- text-reuse is less easy to detect by traditional means, good opportunity to apply computational methods
- Vergil`s influence on Jerome`s writings are “greater than any other secular author” (Hagendahl 1958:305); he was trained by the grammarian Donatus, a Vergilian expert (Mohr 2007:318)
- broad research on reception of Vergil in early Christianity/ by Augustine (Freund 2000, Müller 2003)
- broad research on Vergil in Jerome is missing

Method:

- combine traditional and digital methods ...
- using Tesserae ...
- develop a low-end text-reuse identification instrument

State of research

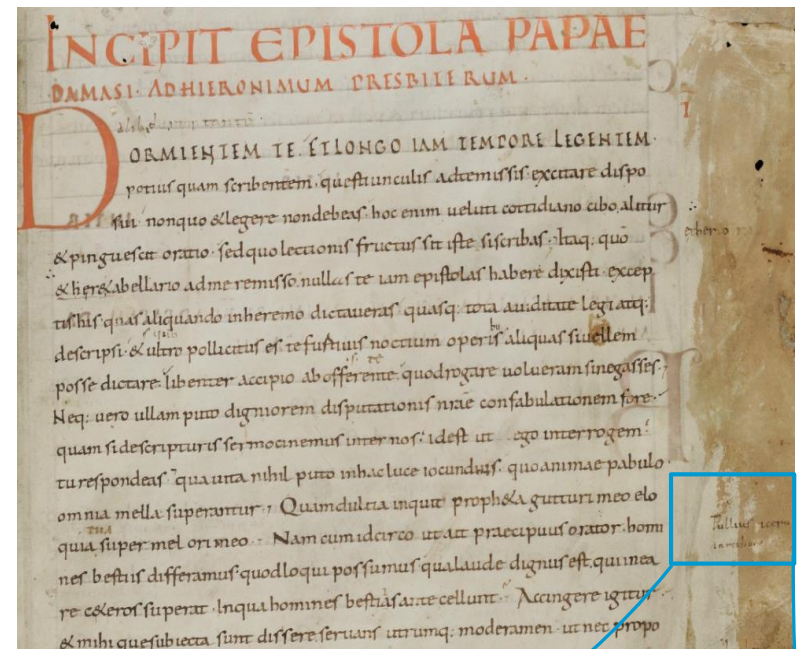
Traditional research (selection):

- fundamental: Hagendahl, Harald (1974, 1958).
- Adkin, Neil: a lot of articles to quotations from different authors, cf. (1997, 2013 Cicero), (1998 Vergil), (2000 Satirists), (2002 Quintilian), (2005 Sallust), (2011a Pliny), (2011b Catullus)
- Luebeck (1872); Kunst (1918); Antin (1960); Godel (1964); Mohr (2007); Cain (2008, 2010, 2013b)
- valuable critical apparatus: Hilberg, Isidor (ed.) (1910, 1912, 1918): CSEL 54-56, Wien. (1996: Vol. 65.2 Indices et addenda)

No digital analyses on Jerome.

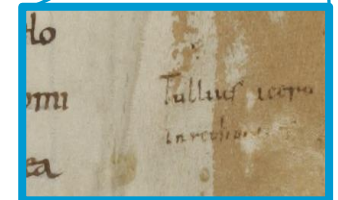
Digital investigation on text-reuse:

e.g. Almas/ Berti (2013) projects site: <http://sites.tufts.edu/perseids>, Bamman/ Crane (2008), Berti et al. (2016), Büchler et al. (2014) projects site: <http://www.etrapp.eu>, Coffee et al. (2012, 2013) projects site: <http://tesserae.caset.buffalo.edu>, Forstall et al. (2015), Mastandrea (2009) projects site: <http://www.mqdq.it>, Scheirer et al. (2016).



Long research tradition:

Marginal gloss: '*ut ait praecipuus orator*' > [Marcus] Tullius Cicero. Codex Sangallensis 159,5 (mid of 9th century), Jer. ep. 35 (CSEL).



Data – Jerome`s letters

- 123 *epistulae*, ~300.000 words
- the *Tesserae* corpus holds a selection of these letters (namely: Wright 1963)
- digitized text of the critical edition (CSEL volumes 54-56, standard printed text version) in the course of the *Open Greek and Latin Project*, Universität Leipzig, TEI-XML
- proofreading:
correct chapter/ paragraph numbers and punctuation marks, missing sentences at the beginning of the letters, redundant characters from OCR process, miss recognitions of: c-e, n-u, 3-8, obviously false words: „virtte“, wrong automatic corrections: „volo“ was frequently corrected to „nolo“, <sic>Komanarum</sic>,
... quo <choice> <sic>mortu</sic>

<corr>mortuus</corr> </choice> us est ..., etc.
(The proofread XML-files are uploaded on the projects GitHub repository again; with many thanks to the project team.)
- transformation of the enhanced XML-files in *Tesserae* conform layout
- incorporation of all letters in the *Tesserae* corpus (special thanks to Neil Coffee and his *Tesserae* team)

Research design (I) – train for correct classification

calibration (1 book of
the *Aeneid*)

A light blue downward-pointing arrow with a white outline, indicating the flow from the calibration step to the evaluation step.

evaluation (another
book of the *Aeneid*)

A light blue downward-pointing arrow with a white outline, indicating the flow from the evaluation step to the use step.

use (all other books)

Research design (IIa) – evaluate binary classification system itself

Main aim:

- correctly classify ‘meaningful’ parallels and ‘not-meaningful’ parallels

Problem:

- not all ‘meaningful’ parallels are known > perfect calibration impossible
- also Jerome`s technique unknown > simulation not possible

Version 1: “traditional philological benchmark”

confusion matrix:

		truth = trad. commentators	
		True	False
estimate	True	TP	FP
	False	FN	TN

True Positive	= found by commentators	maximize
False Positive	= according to comm. no parallels	minimize
False Negative	= true parallels missed by code	minimize
True Negative	= no parallels	

- replicates (old) findings of traditional-manual commentators only (establishes trust in method)

Research design (IIb) – evaluate binary classification system itself

Version 2: expanded truth-term to include new parallels

confusion matrix:

		truth	
		True	False
estimate	True	TP	FP
	False	FN	TN

TP = commentators + algorithm maximize

FP = ballast, no parallels minimize

FN = missed by algorithm

TN = no parallels

- concentrate on first row of matrix only
- distinction of ‘meaningful’ and ‘not-meaningful’* parallels through *close-reading*
- decision if result has to remain positive or has to change to negative
- finds hitherto unknown parallels, too

* the term ‘meaningful’ refers here and elsewhere to the *close-reading* process, in which the decision is made whether a passage with reused text relates to its source text and thereby creates a third meaning, or not. How this additional meaning is created is not covered in this use of the term.

Research design (III) – optimization rules

Optimize filter process including two standards:

Op1: ‘traditional philological benchmark’: do not fall behind *status quo* of research (max. recall)

- find all ‘meaningful’ parallels found by commentators
- do not care about false positives
- very inefficient for *close-reading* (too much results)

Op2: improve classification (max. precision and recall)

- find as many ‘meaningful’ parallels found by commentators as possible
- but reduce also false negatives
- focus optimizing correct classification (either positive or negative)
- provide a more manageable ground for *close-reading*

First steps of analysis

Started with Tesseract search:

- how to find constant filter settings
- apply Op1 and Op2
- customize stoplist

Complement Tesseract search with own python environment:

- working with Tesseract and choosing wide filter settings
(source = vergil.aeneid.part.1, target = jerome.epistulae, unit = phrase, feature = word, stopsize = 0, stbasis = corpus, stopwords =, max_dist = 999, dibasis = freq, cutoff = 0, filter = off)
- replacing further filter settings by own code and criteria (meeting Op1 and Op2)
- derive criteria from traditional search process of finding parallels
- ‘corpus-driven approach’ because of constant interrelated *close-reading*

Filter criteria

stoplists & stoplist-words:

- 9 different consecutive lists, aim: finding best list in respect to Op1 & Op2
- word frequency but, not only function words, also tests with low frequent words
- enhanced through *close-reading* (cf. Jockers 2013:131)

distance:

- two settings: max. 2 or 3 words between two shared words (exclusively) in both texts (near to manual detected more ‘compact’ parallels, next level of word separation to look at)
- shared words also reduced through stoplists
- stoplist words are not counted as in-between words, so the real distance might be higher

shared words:

- at least two shared words

Python code* – data $\triangleq$ Tesseract output

code structure:

- due to the structure of the *Tesseract* output code operates with nested lists with two levels
- manipulated items will be appended on every list (documentation purposes)
- data grows in terms of items not in terms of nested levels

Data normalization (lower case, whitespaces, Ascii/UTF-8)	.lower(), .strip(), .replace(), regex (compiled re.sub())
Data cleaning (standardized orthography (j>i, v>u), punctuation marks and brackets etc.)	.translate()
Apply set of distinct stopwordlists (replace stopword with a single whitespace)	set(), .replace() of key-value tuples
Count distances (= whitespaces) between parallel words for every applied stopwordlist	enumerate(), .count()
Count remaining parallel words for every applied stoplist	slicing, .count()
Retrieval of all relevant results, stoplist separated for comparison (precision/ recall)	enumerate()

*I want to thank PD Dr. Gerold Schneider for his very helpful course “Python für Computerlinguistik” in the winter term 2016/17.

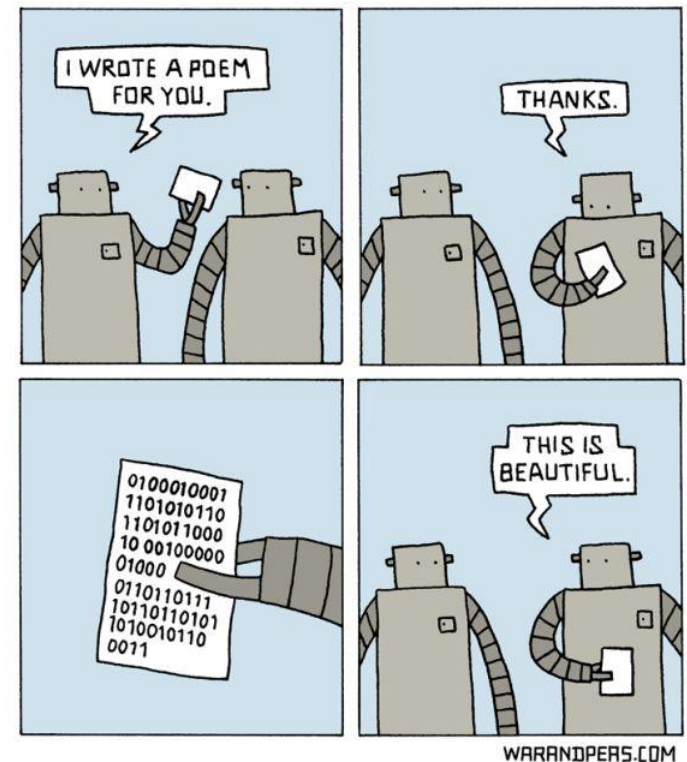
First results and further plans

First results:

- quite acceptable number of estimated positive parallels of Python code for the *Aeneid*
- new meaningful parallels found, e.g.:
jer. ep. 1.5.2 and verg. aen. 1.92: ****solvuntur** **membra****
jer. ep. 64.2.3 and verg. aen. 1.673: ****teneatur** **amore****

Further plans:

- bottleneck of stoplist application in Python
- expand on other works of Vergil but also other authors
- continue *close-reading* of corpus driven results
- syntactic analysis of parallels
- create typology
- draw the line back to the research question
(cf. comic)



References:

Primary source:

- Hilberg, I. (ed.) (1910, 1912, 1918): *Epistulae*. Pars I: I-LXX, CSEL 54; Pars II: LXXI-CXX, CSEL 55; Pars III: CXXI-CLIV, CSEL 56, Wien.
(reprint 1996, incl. Indices et addenda, CSEL. 65.2)
- Wright, F. A. (ed.) (1963): *Select Letters of St. Jerome*, Cambridge.

Secondary literature:

- Adkin, N. (1997): Cicero's 'Orator' and Jerome, *VChr* 51.1, 25–39.
- Adkin, N. (1998): Vergil's 'Georgics' and Jerome, *Epist.* 125,11,3–4, *WJA* 22, 187–198.
- Adkin, N. (2000): Jerome, Seneca, Juvenal, *RBPh* 78, 119–128.
- Adkin, N. (2002): The eleventh Book of Quintilian's *institutio oratoria* and Jerome, *Eos* 89, 315–319.
- Adkin, N. (2005): Hieronymus Sallustianus, *GB* 24, 93–120.
- Adkin, N. (2011a): A New Echo of Pliny the Younger in Jerome?, *Philologus* 155.1, 193–195.
- Adkin, N. (2011b): Catullus in Jerome? Notes on the 'Cohortatoria de paenitentia ad Sabinianum (Epist. 147)', *VChr* 65.4, 408–424.
- Adkin, N. (2013): Cicero's 'Pro Milone' and Jerome, *Euphrosyne* 41, 367–374.
- Almas, B./ Berti, M. (2013): Perseids Collaborative Platform for Annotating Text Re-uses of Fragmentary Authors, in: (Proceedings) 1st International Workshop on Collaborative Annotations in Shared Environment: Metadata, Vocabularies and Techniques in the Digital Humanities, DH-CASE '13, New York, NY.
- Antin, P. (1960): Touches classiques et chrétiennes juxtaposées chez saint Jérôme, *RPh* 34, 58–65.
- Bamman, D./ Crane, G. (2008): The logic and discovery of textual allusion, in: *Proceedings of the Second Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2008)*, Marrakesh.
- Berti, M./ Blackwell C.W./ Daniels, M./ Strickland, S./ Vincent-Dobbins/ K. (2016): Documenting Homeric Text-Reuse in the *Deipnosophistae* of Athenaeus of Naucratis, in: *Digital Approaches and the Ancient World Bulletin of the Institute of Classical Studies* 59, 121–139.
- Büchler, M./ Burns, P./ Müller, M./ Franzini, E./ Franzini, G. (2014): Towards a historical text reuse detection, in: Biemann, C./ Mehler, A. (ed.): *Text Mining. From ontology learning to automated text processing applications*, Heidelberg, 221–238.
- Cain, A. (2008): Liber manet: Pliny, ep. 9.27.2 and Jerome, ep. 130.19.5, *CQ* 58.2, 708–710.
- Cain, A. (2010): Three further echoes of Lactantius in Jerome, *Philologus* 154.1, 88–96.
- Cain, A. (2013a): *Jerome and the Monastic Clergy: A Commentary on Letter 52 to Nepotian, with an Introduction, Text, and Translation*, Leiden/ Boston.
- Cain, A. (2013b): Two Allusions to Terence, Eunuchus 579 in Jerome, *CQ* 63.1, 407–412.

References:

- Coffee, N./ Koenig, J.-P./ Poornima, S./ Ossewaarde, R. (2012): Intertextuality in the Digital Age, TAPhA 142.2, 383-422.
- Coffee, N./ Koenig, J.-P./ Poornima, S./ Forstall, C./ Ossewaarde, R./ Jacobson, S. L. (2013): The Tesserae Project. Intertextual analysis of Latin poetry, Lit Linguist Computing 28.2, 221-228.
- Forstall, C./ Coffee, N./ Buck, T./ Roache, K./ Jacobson, S. (2015): Modeling the scholars: Detecting intertextuality through enhanced word-level n-gram matching, Lit Linguist Computing 30.4, 503-515.
- Freund, S. (2000): Vergil im frühen Christentum. Untersuchungen zu den Vergilzitate bei Tertullian, Minucius Felix, Novatian, Cyprian und Arnobius, Paderborn.
- Fürst, A. (2016): Hieronymus. Askese und Wissenschaft in der Spätantike, Freiburg/ Basel/ Wien.
- Godel, R. (1964): Réminiscences de poètes profanes dans les lettres de St-Jérôme, MH 21.1, 65-70.
- Hagendahl, H. (1958): Latin Fathers and the Classics. A study on the Apologists, Jerome and other Christian writers, Göteborg.
- Hagendahl, H. (1974): Jerome and the Latin Classics, VChr 28.3, 216-227.
- Jockers, M. L. (2013): Macroanalysis. Digital Methods and Literary History, Urbana Ill.
- Kunst, K. (1918): De S. Hieronymi studiis ciceronianis. Diss., Wien/ Leipzig.
- Luebeck, E. (1872): Hieronymus quos nouerit scriptores et ex quibus hauserit, Lipsiae.
- Mastandrea, P. (2009): Gli archivi elettronici di 'Musisque deoque'. Ricerca intertestuale e cernita fra varianti antiche, (con qualche ripensamento sulla tradizione indiretta dei poeti latini), in: Zurli, L./ Mastandrea, P. (ed.): In Poesia latina. Nuova e-filologia. Opportunità per l'editore e per l'interprete, Rom, 41-72.
- Mohr, A. (2007): Jerome, Virgil, and the captive maiden. The attitude of Jerome to classical literature, in: Scourfield, J. H. D. (ed.): Texts and culture in late antiquity. Inheritance, authority, and change, Swansea, 299-322.
- Müller, G. A. (2003): Formen und Funktionen der Vergilzitate bei Augustin von Hippo. Formen und Funktionen der Zitate und Anspielungen, Paderborn.
- Scheirer, W./ Forstall, C./ Coffee, N. (2016): The sense of a connection: Automatic tracing of intertextuality by meaning, Lit Linguist Computing 31.1, 204-217.

Illustrations:

Dürer, Albrecht (1514): 'Der heilige Hieronymus im Gehäuse', Source: Wikimedia Commons, public domain.

St. Gallen, Stiftsbibliothek, Cod. Sang. 159, p. 5. (<http://www.e-codices.unifr.ch/de/csg/0159/5>)

Comic: <https://warandpeas.com/>

Thank
You!

cursim quasi de prato pulcherrimo (...)

paruos flores carpsisse

(Jer. ep. 130,9,1)

Stimuli for the discussion

Python coding: bottleneck concerning time efficiency when applying different stoplists

Corpus-based (frequency only) or corpus-driven (dominant word categories)
compilation of stoplist words?

Implicit intertextuality theory affects operationalization (circle)

Does the quality of the digital text affect the results of *text-reuse* analyses (in a relevant way)?

Mixing methods: How to get back to the original question? Results of digital analysis
kind of different “ontological” status.