

Citations and Annotations in Classics: Old Problems and New Perspectives

Matteo Romanello (KCL, DAI)
Michele Pasin (Nature)

DH-CASE '13 @ DocEng – Firenze, 10 September 2013

Section 1

HuCit in Context

My Research

PhD in Digital Humanities Research

- ▶ Willard McCarty (KCL, DDH)
- ▶ Shalom Lappin (KCL, Phil. Dept.)

In a Nutshell

- ▶ study on how citations to ancient texts contained in modern publications can be captured and exploited in order to offer new ways of studying texts in Classics (e.g. text reception, intertextuality, etc.)

What are Canonical Citations 1

It is not, in general, permissible to alter a text to suit historical or geographical fact, unless there is something in the language itself which also demands a change: a principle which justifies, for example, the insertion of a negative in *Thuc. VIII. 101. 1* and the obelizing of *νέωρεος* in *Ar. Const. Ath. 24. 3*. But there are also cases in which the error is certain and the author most unlikely to have committed it; if both these conditions are present, we may suspect the text. It is generally agreed that these last two conditions are present in this passage of Herodotus; but it has not been noticed (I believe) that the first is present also. Blakesley indeed accepted the text at its face value, and considered it a triumphant proof that the boasts of Athens about freeing the Greeks of Asia from the barbarian were all false—a fact that seems to have escaped her contemporary critics: and others have toyed with this explanation. But if the direct statement of *Thucydides (VIII. 5. 5)* and the terms of the treaty between Persia and Sparta (*VIII. 18*) are not sufficient to refute it, consider these two facts: (i.) Erythrae had an Athenian garrison, from about 460, and the Athenian commander had some control even over her internal affairs; she could not restore exiles hostile to Athens: did she pay tribute to Persia? (ii.) In the quarrel between Miletus and Samos in 441, it is Miletus, the city on the mainland, that appeals to Athens; and Pissuthnes and the Samian oligarchs prepare to march against her. Why, if the Persians were strong enough to exact tribute? Macan argues that many new cities became tributary to Athens for the first time after Eurymedon, and many were lost again and became tributary to Persia after the Egyptian disaster. But neither of these changes affected Ionia; and if they had, it would make Herodotus' statement even more untrue, for he asserts that there was no change. Macan also adduces *Thuc. I. 138. 5*, the gift by Xerxes to Themistocles of Magnesia, Myus, and Lampsacus. But note that it was only in Magnesia that he

What are Canonical Citations 2

APh 75-06697

In **Statius' « Achilleid » (2, 96-102)** Achilles describes [...]. The portrayal of angry warriors in Roman epic is effected for the most part not by direct descriptions but indirectly, by similes of wild beasts (e.g. **Vergil, Aen. 12, 101-109 ; Lucan 1, 204-212 ; Statius, Th. 12, 736-740 ; Silius 5, 306-315**). These similes may be compared to two passages from Statius (**Th. 1, 395-433 and 8, 383-394**) that portray the onset of anger in direct narrative. Analysis of these passages demonstrates that the concept of « ira » in epic takes its moral aspect from the context.

Canonical Citations

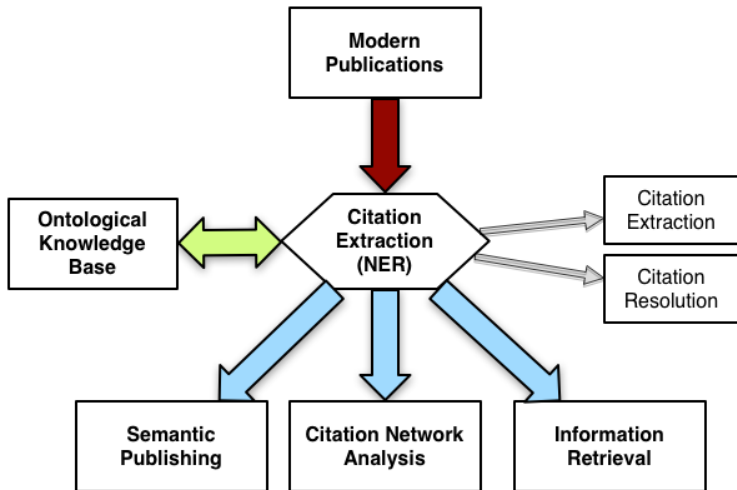
Characteristics

- ▶ refer to (one of) the very object of research, = texts
- ▶ references to text
 - ▶ **precise**
 - ▶ **interoperable**
- ▶ work as resolvable pointers

Challenges

- ▶ ambiguity
- ▶ variation (pub. venue, time)
- ▶ context, underspecification

the big picture



Extracting Citations and their Meaning

1 Identify Citation Components (Named Entities Extraction)

In Life of Lysander (22) Plutarch repeats part of the
in Xenophon (Hell. 3.3.1-4). Xenophon relates
the flight from the chamber is related in the form of a dialogue between
Leotychidas replies that his mother, who knows, says,

machine-learning based NE tagger

Secondary Literature

- Thuc., 109, 4; □ Samuel K. Eddy, "Some Irregular Amounts of Athenian Tribute" in *The American Journal of Philology* vol. 94 pp. 47-70 (read on JSTOR)
- discussed
□ Thuc. (2)
○ Thuc., 1, 104, 1-3;
○ Thuc., 1, 115, 4.
- Thuc. 1, 109, 2-3; □ Thomas Kelly, "Argive Foreign Policy in the Fifth Century B.C." in *Classical Philology* vol. 69 pp. 81-99 (read on JSTOR)
- discussed
□ Thuc. (64)
□ Diod. (9)
□ Ar. (1)
□ Hdt. (11)
○ Hdt. 7, 151-52.
○ Hdt. 148-52, 9, 12.
○ Hdt. 7.
○ Hdt. 221-29, 3.
○ Hdt. 6, 75-84.
○ Hdt. 6, 92.

2 Determine Relations Between Them (Relation Extraction)

In Life of Lysander (22) Plutarch repeats part of the account.
in Xenophon (Hell. 3.3.1-4). Xenophon relates
the flight from the chamber is related in the form of a dialogue between
his son. Leotychidas replies that his mother, who knows, says,

rule-based parsing

3 Assign Unique Identifier To Each Citation (NE Disambiguation)

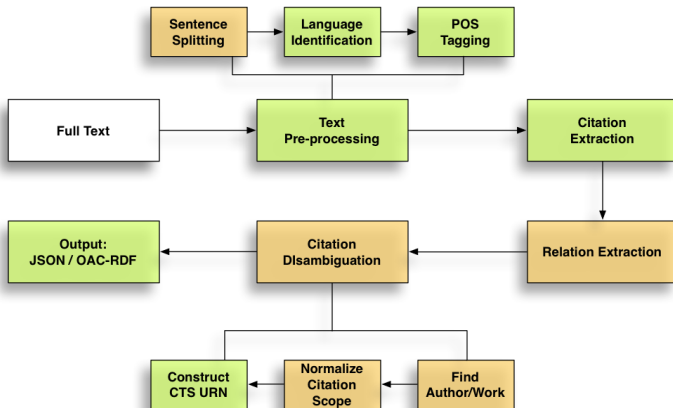
tears to declare him his son in the
ake. In Life of Lysander (22) Plutarch
ence in Xenophon (Hell. 3.3.1-4). Xenophon relates that after the death of Agis, Leotychidas
the flight from the chamber is related in the form of a dialogue between Agisilaus and Leotychidas. Agis
not his son. Leotychidas replies that his mother, who knows, says Agis was his father. Agisilaus retors

REFAUWORK → Scope → REFScope 23-34
"Hell." → "3.3.1-4".
Note: urn:cts:greekLit:tlg0032.tlg001:3.3.1-3.3.4

rule-based parsing

Annotation UI << <http://brat.nlpplab.org/>

NLP Pipeline



Related Work: CITE/CTS & CWKB OpenURL

CWKB OpenURL (CornellU)

- ▶ `http://cwkb.org/resolver?ctx_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:canonical_cit&rft.workid=http://cwkb.org/workid/tlg:0003.001&rft.au=Thucydides&rft.slevel1=4&rft.slevel2=8&rft.slevel3=3&rft.elevel3=9&rft_id=info:sid/aph%22`

Cite/CTS (HomerMultiText)

- ▶ `http://www.perseus.tufts.edu/hopper/CTS?request=GetPassage=urn:cts:greekLit:tlg0003.tlg001:4.8.3-4.8.9`

CTS Protocol

Goal

- ▶ repository of TEI-encoded texts
- ▶ identifiable citation scheme

Methods

- ▶ GetCapabilities
- ▶ GetPassage + CTS URN
- ▶ GetPassagePlus + CTS URN
- ▶ GetValidReff

Canonical Text Service (CTS)

CTS URNs

- ▶ `urn:cts:greekLit:tlg0003` (author = Thucydides)
- ▶ `urn:cts:greekLit:tlg0003.tlg001` (work = Histories)
- ▶ `urn:cts:greekLit:tlg0003.tlg001:4.8.3`
- ▶
`urn:cts:greekLit:tlg0003.tlg001.perseus-grc1:4.8.3#<str>[idx]`
- ▶ `urn:cts:greekLit:tlg0003.tlg001.perseus-eng1:4.8.3-4.8.9`

Section 2

HuCit: why and how?

Citation Ontologies

Two main approaches

1. Citation as *scholarly act* (relation)
2. Citation as textual phenomenon (object, pointer)

Terminological Ambiguity

► Reference

- identifies uniquely a publication
- provides enough information to retrieve it

► Citation

- occurs in the body of a document
- points to a reference at the end of it

Citations in Existing Ontologies

As performative entity

- ▶ `bibo:cites`
- ▶ `akt-cites:publication-reference`
- ▶ `cito:cites`
 - ▶ `cito:cites-as-evidence`

As textual phenomenon

- ▶ `biro:BibliographicReference`
- ▶ `biro:ReferenceList`
- ▶ `doco:BibliographicReferenceList`
- ▶ `deo:BibliographicReference`

Methodology

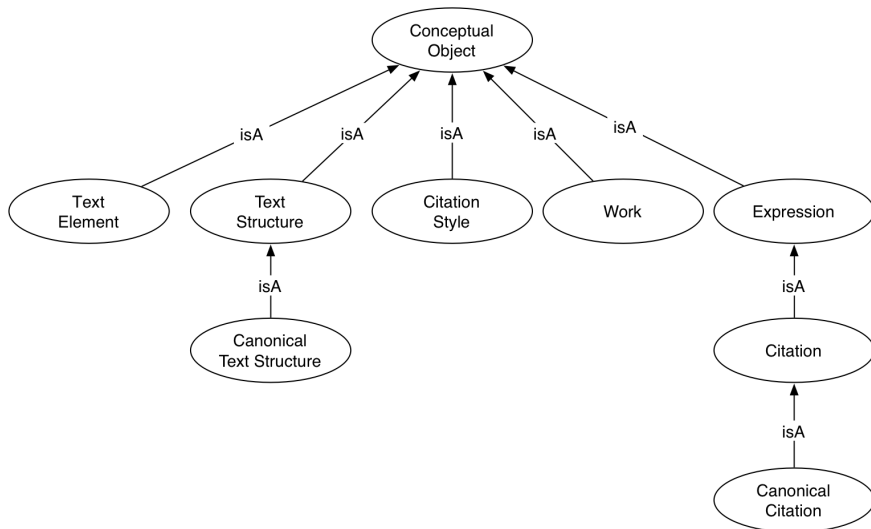
Ontology of Representations (Mizogouchi-Pasin)

- ▶ *representations*: information- or content-bearing objects
- ▶ representation has *form* and *content*
- ▶ **form** and **content** of canonical citations?

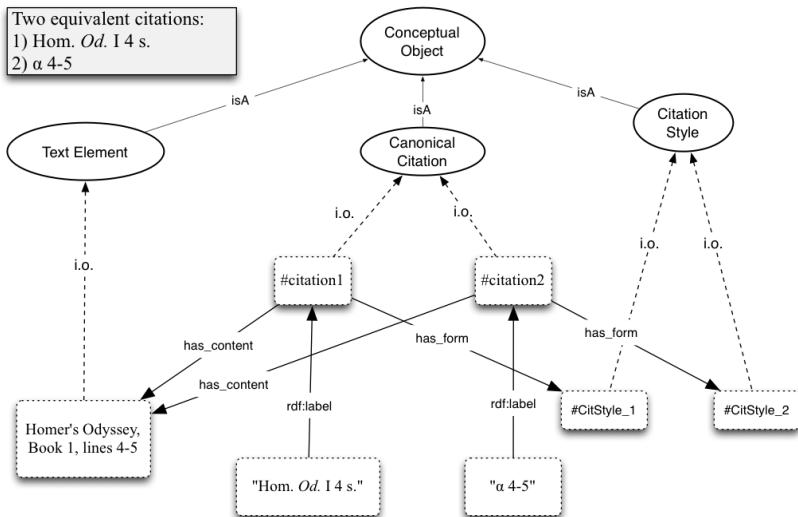
Reuse and Extension of

- ▶ CIDOC-CRM
- ▶ FRBRoo

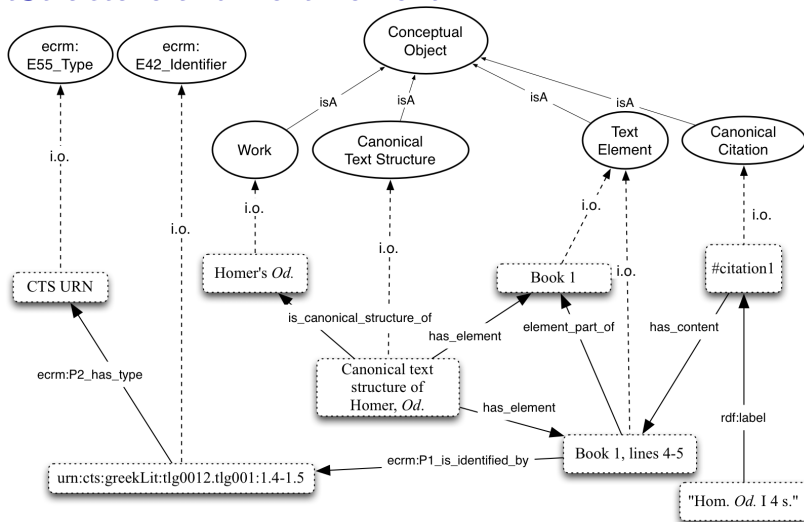
HuCit overview



Form and Content of a (Canonical) Citation



TextStructure and TextElement



Section 3

HuCit in Action

1 Identify Citation Components (Named Entities Extraction)

In Life of Lysander (22) Plutarch repeats part of the

in Xenophon (Hell. 3.3.1-4). Xenophon relates the flight from the chamber is related in the form of a dialogue between Leotychidas replies that his mother, who knows, says:

machine-learning based NE tagger



2 Determine Relations Between Them (Relation Extraction)

In Life of Lysander (22) Plutarch repeats part of the account.

in Xenophon (Hell. 3.3.1-4). Xenophon relates the flight from the chamber is related in the form of a dialogue between his son. Leotychidas replies that his mother, who knows, says:

rule-based parsing



3 Assign Unique Identifier To Each Citation (NE Disambiguation)

tears to declare him his son in the

ake. In Life of Lysander (22) Plutarch

urce in Xenophon (Hell. 3.3.1-4). Xenophon relates that after the death of Agis, Leotychidas the flight from the chamber is related in the form of a dialogue between Agesilaus and Leotychidas. Agis not his son. Leotychidas replies that his mother, who knows, says Agis was his father. Agesilaus retors

REFAUWORK → Scope → REFScope ID:84
"Hell." → "3.3.1-4)."
Note: urn:cts:greekLit:tlg0032.tlg001:3.3.1-3.3.4

rule-based parsing

Secondary Literature

- Thuc., 109, 4; [Samuel K. Eddy, "Some Irregular Amounts of Athenian Tribute" in The American Journal of Philology vol. 94 pp. 47-70 [read on JSTOR]
- Also cited:
 • Thuc. (2)
 • Thuc., 1, 104, 1-3;
 • Thuc., 1, 115, 4.
- Thuc. 1. 109. 2-3; [Thomas Kelly, "Argive Foreign Policy in the Fifth Century B.C." in Classical Philology vol. 69 pp. 81-99 [read on JSTOR]
- Also cited:
 • Thuc. (64)
 • Diod. (9)
 • Ar. (1)
 • Hdt. (11)
 • Hdt. 7. 151-52.
 • Hdt. 149-52, 9. 12,
 • Hdt. 7.
 • Hdt. 221-29. 3.
 • Hdt. 6. 75-84.
 • Hdt. 6. 92



Annotation UI << <http://brat.nlplab.org/>

Reasoning: valid citation

- ▶ input: “Hom. *Il.* 1.1-10”
- ▶ which text(s) can be abbreviated by “Hom. *Il.*”?
 - ▶ *Iliad*
- ▶ how many levels has the TextStructure of the *Iliad*?
 - ▶ book,line
 - ▶ citation has 2 *levels* as well
- ▶ does *Iliad* book 1 contain lines 1 to 10?
 - ▶ yes
- ▶ what is *Iliad*'s CTS URN?
 - ▶ urn:cts:greekLit:tlg0012.tlg001
- ▶ output: urn:cts:greekLit:tlg0012.tlg001:1.1-1.10

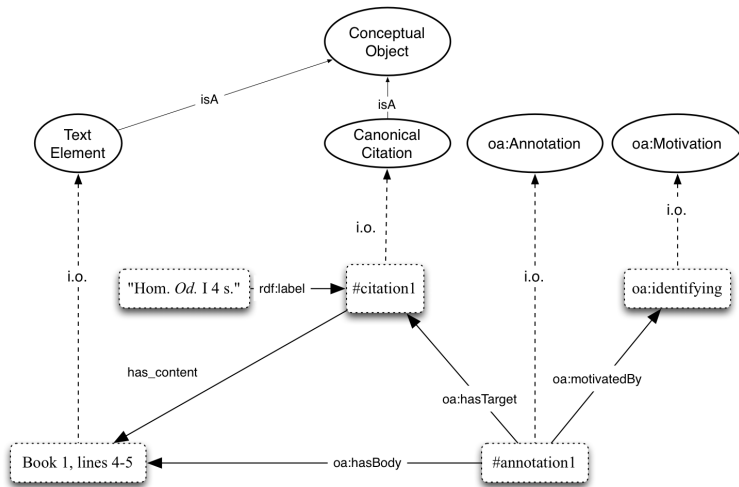
Reasoning: invalid citation

- ▶ input: “Hom. *Il.* 1.10.1”
- ▶ which text(s) can be abbreviated by “Hom. *Il.*”?
 - ▶ *Iliad*
- ▶ how many levels has the TextStructure of the *Iliad*?
 - ▶ book,line
- ▶ error: citation has 3 levels

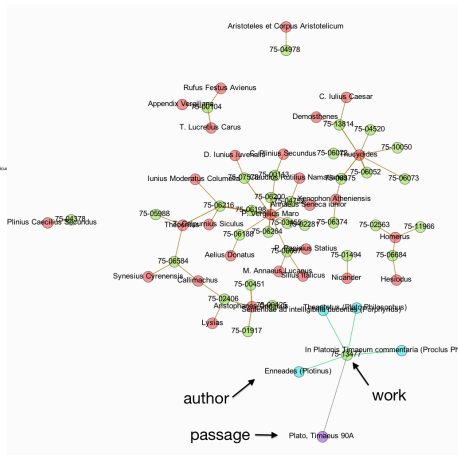
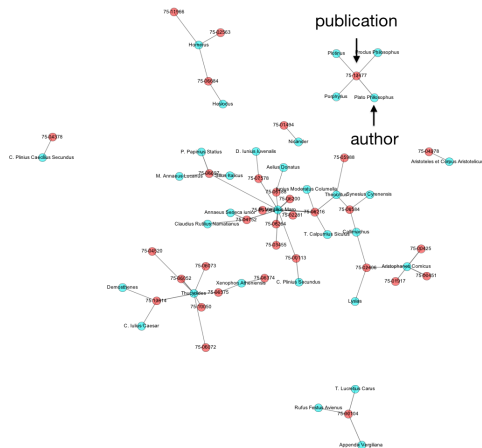
Reasoning: disambiguation

- ▶ input: “Th. 1.33”
- ▶ which text(s) can be abbreviated by *Th.*?”
 - ▶ Thucydides’ *Histories*
 - ▶ Theocritus’ *Idylls*
- ▶ levels of the canonical structure:
 - ▶ *Histories* : book, chapter, line (3)
 - ▶ *Idylls* : idyll, line (2)
- ▶ does *Histories* book 1 contain chapter 33?
 - ▶ yes
- ▶ does Theocritus’ *Idyll* 1 contain line 33?
 - ▶ yes
- ▶ consider broader context:
 - ▶ “[...] as attested in the famous **idyll** (Th. 1.33) [...]”

HuCCit + OAC (data publishing)



Exporting from HuCit (data analysis)



Section 4

Wrap-up

HuCit's Contribution

- ▶ formal model of citations in Classics
- ▶ CTS
 - ▶ way of querying *knowledge* contained in CTS repositories
 - ▶ semantics of URNs are defined explicitly and machine-readable
- ▶ way of publishing extracted citations as LOD

Work in Progress

- ▶ ontology population via Perseus' CTS API
- ▶ implementation of reasoning in the NLP pipeline

Thanks for your attention! Comments, questions?

`matteo.romanello@kcl.ac.uk`

`http://www.essepuntato.it/lode/owlapi/http://purl.org/net/hucit`

`https://bitbucket.org/56k/hucit/`