An Ontological View of Canonical Citations

Matteo Romanello[1] Michele Pasin[1]

[1]Department of Digital Humanities
King's College London

Digital Humanities conference 2011
Stanford, California
21 June 2011

Project Context

- PhD at the DDH, KCL
- discussion started during an internal Knowledge Representation seminar
- HuCit: ontology for the citations in the Humanities
- part of a domain ontology for Classics
- application: knowledge base to support the automatic extraction of citations from text

Related Work

Ontology Engineering

- CiTO, Citation Typing Ontology
- SWAN Ecosystem, esp. Scientific Discourse Ontology
- OpenCyc

Canonical Citations

- Canonical Text Services (CTS), Harvard Univ.
- Classical Works Knowledge Base (CWKB), Cornell Univ.

Canonical Citations

Motivation

- **precision** of references to texts
- **persistency** of citations
- **interoperability** of citations across editions

Characteristics

- are not tight to a specific edition of a text (loosely coupled)
- canonical: result of an agreement within a specific community of practice

Examples

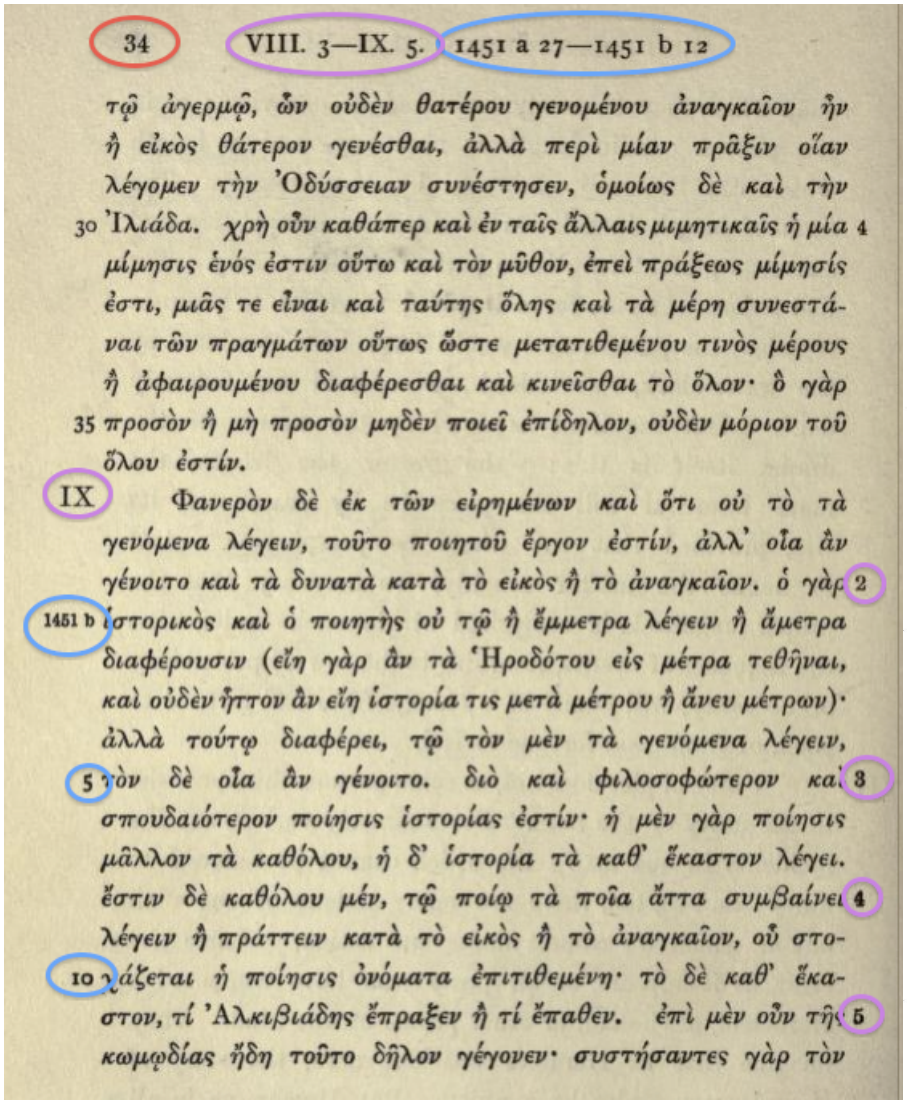
- Hom. *Il.* 1.1
- Athen. *Deipn,* X 412a
- Arist. *Poetics* 1451a35-b6

Citing Aristotle

- Bekker numbers
  - derived from his edition (1831-1870) of Aristotle's works
  - used by **all** scholarly editions
- citation structure:
  - 1451a35
  - page / column / line
  - / [0-9]{4}[a-z]{1}[0-9]+ /
- alternative citation structure:
  - eg: VIII 4 - IX 3
  - medieval system
  - book / chapter / sentence
- Similar systems:
  - Stephanus page (Plutarch)
  - Casaubon page (Athenaeus)

`http://en.wikipedia.org/wiki/Bekker_numbers`

Citations in a print environment



34 VIII. 3—IX. 5. 1451 a 27—1451 b 12

τῷ ἀγερμῷ, ὧν οὐδὲν θατέρου γενομένου ἀναγκαῖον ἦν ἢ εἰκὸς θάτερον γενέσθαι, ἀλλὰ περὶ μίαν πρᾶξιν οἵαν λέγομεν τὴν Ὀδύσσειαν συνέστησεν, ὁμοίως δὲ καὶ τὴν Ἰλιάδα. χρὴ οὖν καθάπερ καὶ ἐν ταῖς ἄλλαις μιμητικαῖς ἡ μία μίμησις ἑνός ἐστιν οὕτω καὶ τὸν μῦθον, ἐπεὶ πράξεως μίμησίς ἐστι, μιᾶς τε εἶναι καὶ ταύτης ὅλης καὶ τὰ μέρη συνεστάναι τῶν πραγμάτων οὕτως ὥστε μετατιθεμένου τινὸς μέρους ἢ ἀφαιρουμένου διαφέρεσθαι καὶ κινεῖσθαι τὸ ὅλον· ὃ γὰρ προσὸν ἢ μὴ προσὸν μηδὲν ποιεῖ ἐπίδηλον, οὐδὲν μόριον τοῦ ὅλου ἐστίν.

IX Φανερὸν δὲ ἐκ τῶν εἰρημένων καὶ ὅτι οὐ τὸ τὰ γενόμενα λέγειν, τοῦτο ποιητοῦ ἔργον ἐστίν, ἀλλ' οἷα ἂν γένοιτο καὶ τὰ δυνατὰ κατὰ τὸ εἰκὸς ἢ τὸ ἀναγκαῖον. ὁ γὰρ ἱστορικὸς καὶ ὁ ποιητὴς οὐ τῷ ἢ ἔμμετρα λέγειν ἢ ἄμετρα διαφέρουσιν (εἴη γὰρ ἂν τὰ Ἡροδότου εἰς μέτρα τεθῆναι, καὶ οὐδὲν ἧττον ἂν εἴη ἱστορία τις μετὰ μέτρου ἢ ἄνευ μέτρων)· ἀλλὰ τούτῳ διαφέρει, τῷ τὸν μὲν τὰ γενόμενα λέγειν, τὸν δὲ οἷα ἂν γένοιτο. διὸ καὶ φιλοσοφώτερον καὶ σπουδαιότερον ποίησις ἱστορίας ἐστίν· ἡ μὲν γὰρ ποίησις μᾶλλον τὰ καθόλου, ἡ δ' ἱστορία τὰ καθ' ἕκαστον λέγει. ἔστιν δὲ καθόλου μέν, τῷ ποίῳ τὰ ποῖα ἄττα συμβαίνει λέγειν ἢ πράττειν κατὰ τὸ εἰκὸς ἢ τὸ ἀναγκαῖον, οὗ στοχάζεται ἡ ποίησις ὀνόματα ἐπιτιθεμένη· τὸ δὲ καθ' ἕκαστον, τί Ἀλκιβιάδης ἔπραξεν ἢ τί ἔπαθεν. ἐπὶ μὲν οὖν τῆς κωμῳδίας ἤδη τοῦτο δῆλον γέγονεν· συστήσαντες γὰρ τὸν

184 ΠΕΡΙ ΣΟΦΙΣΤΙΚΩΝ ΕΛΕΓΧΩΝ.

ἑκάτεροι τὸν ἀλλήλων λόγον. διόπερ ταχεῖα μὲν ἄτεχνος δ' ἂν ἡ διδασκαλία τοῖς μανθάνουσι παρ' αὐτῶν· οὐ γὰρ τέχνην ἀλλὰ τὰ ἀπὸ τῆς τέχνης διδόντες παιδεύειν ὑπελάμβανον, ὥσπερ ἂν εἴ τις ἐπιστήμην φάσκων παραδώσειν ἐπὶ τὸ μηδὲν πονεῖν τὸν πόδα, εἶτα σκυτοτομικὴν μὲν μὴ διδάσκοι, μηδ' ὅθεν δυνήσεται πορίζεσθαι τὰ τοιαῦτα, δοίη δὲ πολλὰ γένη παντοδαπῶν ὑποδημάτων· οὗτος γὰρ βεβοήθηκε μὲν πρὸς τὴν χρείαν, τέχνην δ' οὐ παρέδωκεν. καὶ περὶ μὲν τῶν ῥητορικῶν ὑπῆρχε πολλὰ καὶ παλαιὰ τὰ λεγόμενα, περὶ δὲ τῆ συλλογίζεσθαι παντελῶς οὐδὲν εἶχομεν πρότερον ἄλλο λέγειν, ἀλλ' ἢ τριβῇ ζητοῦντες πολὺν χρόνον ἐπονοῦμεν. εἰ δὲ φαίνεται θεασαμένοις ὑμῖν ὡς ἐκ τούτων ἐξ ἀρχῆς ὑπαρχόντων ἔχειν ἡ μέθοδος ἱκανῶς παρὰ τὰς ἄλλας πραγματείας τὰς ἐκ παραδόσεως πεζευμένας, λοιπὸν ἂν εἴη πάντων ὑμῶν ἢ τῶν ἠκροαμένων ἔργον τοῖς μὲν παραλελειμμένοις τῆς μεθόδου συγγνώμην τοῖς δ' εὑρημένοις πολλὴν ἔχειν χάριν.

ΦΥΣΙΚΗΣ ΑΚΡΟΑΣΕΩΣ Α.

Ἐπειδὴ τὸ εἰδέναι καὶ τὸ ἐπίστασθαι συμβαίνει περὶ πάντας τὰς μεθόδους, ὧν εἰσὶν ἀρχαὶ ἢ αἴτια ἢ στοιχεῖα, ἐκ τοῦ ταῦτα γνωρίζειν (τότε γὰρ οἰόμεθα γινώσκειν ἕκαστον, ὅταν τὰ αἴτια γνωρίσωμεν τὰ πρῶτα καὶ τὰς ἀρχὰς τὰς πρώτας καὶ μέχρι τῶν στοιχείων), δῆλον ὅτι καὶ τῆς περὶ φύσεως ἐπιστήμης πειρατέον διορίσασθαι πρῶτον τὰ περὶ τὰς ἀρχάς. πέφυκε δὲ ἐκ τῶν γνωριμωτέρων ἡμῖν ἡ ὁδὸς καὶ σαφεστέρων ἐπὶ τὰ σαφέστερα τῇ φύσει καὶ γνωριμώτερα· οὐ γὰρ ταὐτὰ ἡμῖν τε γνώριμα καὶ ἁπλῶς. διόπερ ἀνάγκη τὸν τρόπον τοῦτον προάγειν ἐκ τῶν ἀσαφεστέρων μὲν τῇ φύσει ἡμῖν δὲ σαφεστέρων ἐπὶ τὰ σαφέστερα τῇ φύσει καὶ γνωριμώτερα. ἓξ δ' ἡμῖν πρῶτον ὅλα καὶ σαφῆ τὰ συγκεχυμένα μᾶλλον· ὕστερον δ' ἐκ τούτων γίνεται γνώριμα τὰ στοιχεῖα καὶ αἱ ἀρχαὶ διαιροῦσι ταῦτα. διὸ ἐκ τῶν καθόλου ἐπὶ τὰ καθ' ἕκαστα δεῖ προϊέναι. τὸ γὰρ ὅλον κατὰ τὴν αἴσθησιν γνωριμώτερον, τὸ δὲ καθόλου ὅλον τί ἐστιν· πολλὰ γὰρ περιλαμβάνει ὡς μέρη τὸ καθόλου. πέφυκε δὲ ταὐτὸ τοῦτο τρόπον τινὰ καὶ τὰ ὀνόματα πρὸς τὸν λόγον· ὅλον γάρ τι καὶ ἀδιορίστως σημαίνει, οἷον ὁ κύκλος· ὁ δὲ ὁρισμὸς αὐτὸ διαιρεῖ εἰς τὰ καθ' ἕκαστα. καὶ τὰ παιδία τὸ μὲν πρῶτον προσαγορεύει πάντας τοὺς ἄνδρας πατέρας καὶ μητέρας τὰς γυναῖκας, ὕστερον δὲ διορίζει τοῦτον ἑκάτερον.

Ἀνάγκη δ' ἤτοι μίαν εἶναι τὴν ἀρχὴν ἢ πλείους, καὶ εἰ μίαν, ἢ κινητὴν ἢ ἀκίνητον, ὡς φησι Παρμενίδης καὶ Μέλισσος, ἢ κινουμένην, ὥσπερ οἱ φυσικοί, οἱ μὲν ἀέρα φάσκοντες εἶναι οἱ δ' ὕδωρ τὴν πρώτην ἀρχήν· εἰ δὲ πλείους, ἢ πεπερασμένας ἢ ἀπείρους, καὶ εἰ πεπερασμένας πλείους δὲ μιᾶς, ἢ δύο ἢ τρεῖς ἢ τέτταρας ἢ ἄλλον τινὰ ἀριθμόν, καὶ εἰ ἀπείρους, ἢ ὥτως ὥσπερ Δημόκριτος, τὸ γένος ἕν, σχήματι δὲ ἢ εἴδει διαφέρουσας, ἢ καὶ ἐναντίας. ὁμοίως δὲ ζητοῦσι καὶ οἱ τὰ ὄντα ζητοῦντες τόσα· ἐξ ὧν γὰρ τὰ ὄντα ἐστί, πρῶτον ζητοῦσι ταῦτα πότερον ἓν ἢ πολλά, καὶ εἰ πολλά, πεπερασμένα ἢ ἄπειρα, ὥστε τὴν αὐτὴν γνωριμώτερον, τὸ δὲ καθόλου ὅλον τί ἐστιν· ἀρχὴν καὶ τὸ στοιχεῖον ζητοῦσι πότερον ἓν ἢ πολλά. τὸ μὲν ὂν εἰ ἓν καὶ ἀκίνητον τὸ ὂν σκοπεῖν ὑ περὶ φύσεως ἐστι σκο-

3. τῆς om *D.* ‖ 6. δυνήσονται *BD.*
3. ἡμῖν *ABD.* ‖ 6. ἡμῖν *ABD.*
Codices *EFI.* Tit. ἀκροάσεως] ἀκροάσεως ἢ περὶ ἀρχῶν *F.* om titulum *I.*
15. πρῶτον διορίσασθαι *F.* ‖ 16. ἡ om *I.* ‖ 18. τὰ αὐτὰ *F.* ‖ 19. τοῦτον τὸν τρόπον *F.* ‖ μὲν om *FI.* ‖ 20. τῇ om *I.* ‖ τῇ φύσει om pr *E.* ‖ 21. πρῶτον] τὸ πρῶτον *FI.* ‖ 22. συγκρίμενα *F.* ‖ γίνεται post γνώριμα *I*, om *F.* ‖ 24. ἐπὶ] εἰς *E.* ‖ 26. ὡς] ὥσπερ *E.*
2. ἀδιόριστον *I* et corr *E.* ‖ 12. τὸ] δὲ τὸ *I.* ‖ 13. μὲν om *F.* ‖ ὑπολαμβάνει *I*, ὑπολαμβάνει προσαγορεύει πατέρας *E.* ‖ 14. δὲ om *F.* ‖ 16. ὥσπερ *FI.* ‖ φασὶ *I.* ‖ 21. δὲ om *F.* ‖ καὶ om *I.* ‖ 23. ζητοῦσι πρῶτον *FI.* ‖ 26. ὂν] ἓν *I.*

Butcher's edition (1922) p. 34 and Bekker's edition (1831) p. 184

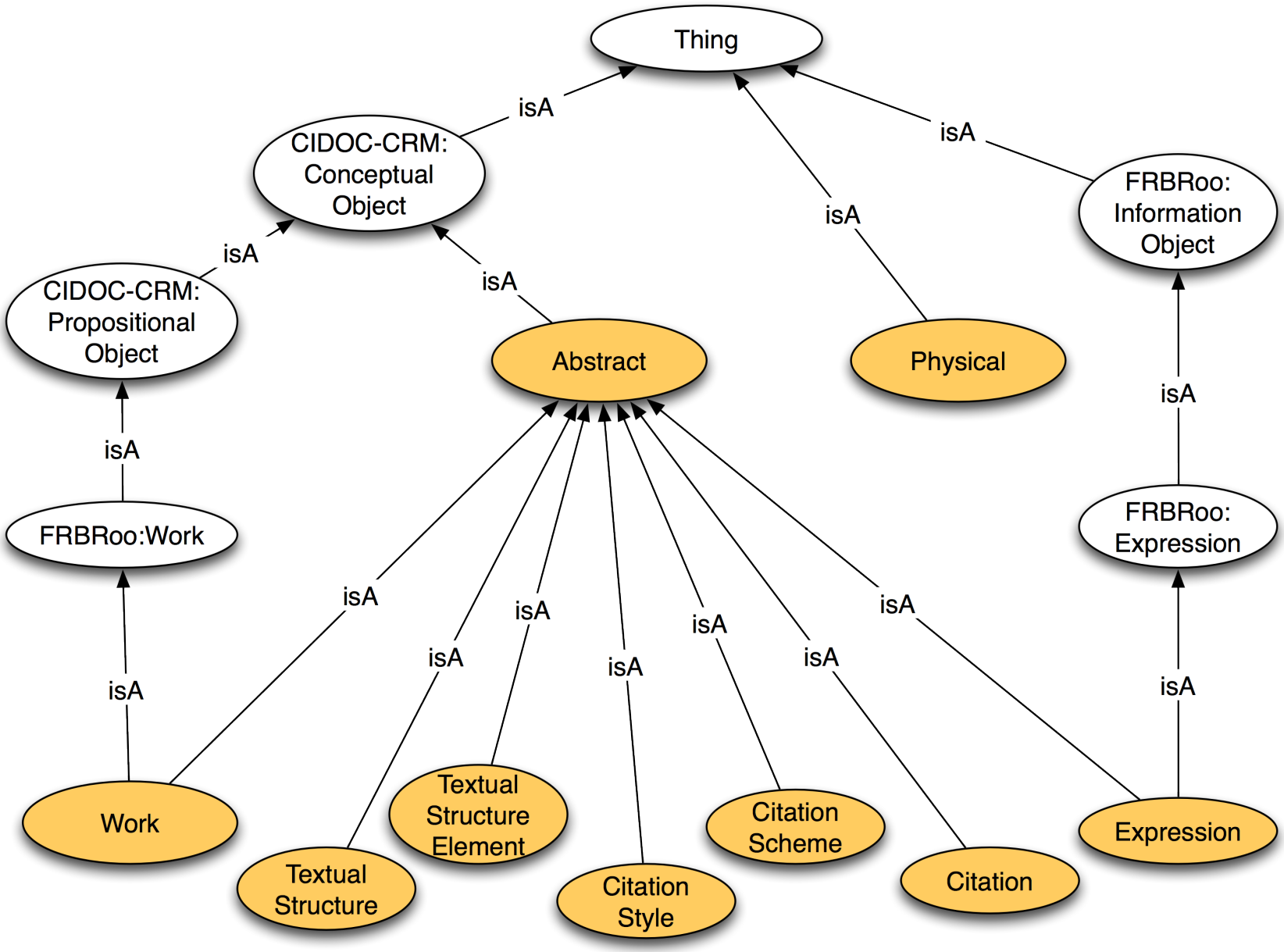
FRBR in a nutshell

- FRBR: Functional Requirements for Bibliographic Records
- **Work**: the abstract content of a Work (substance is concepts)
- **Expression**: the text of a Work (substance is signs)
- **Manifestation**: the physical carrier of an Expression

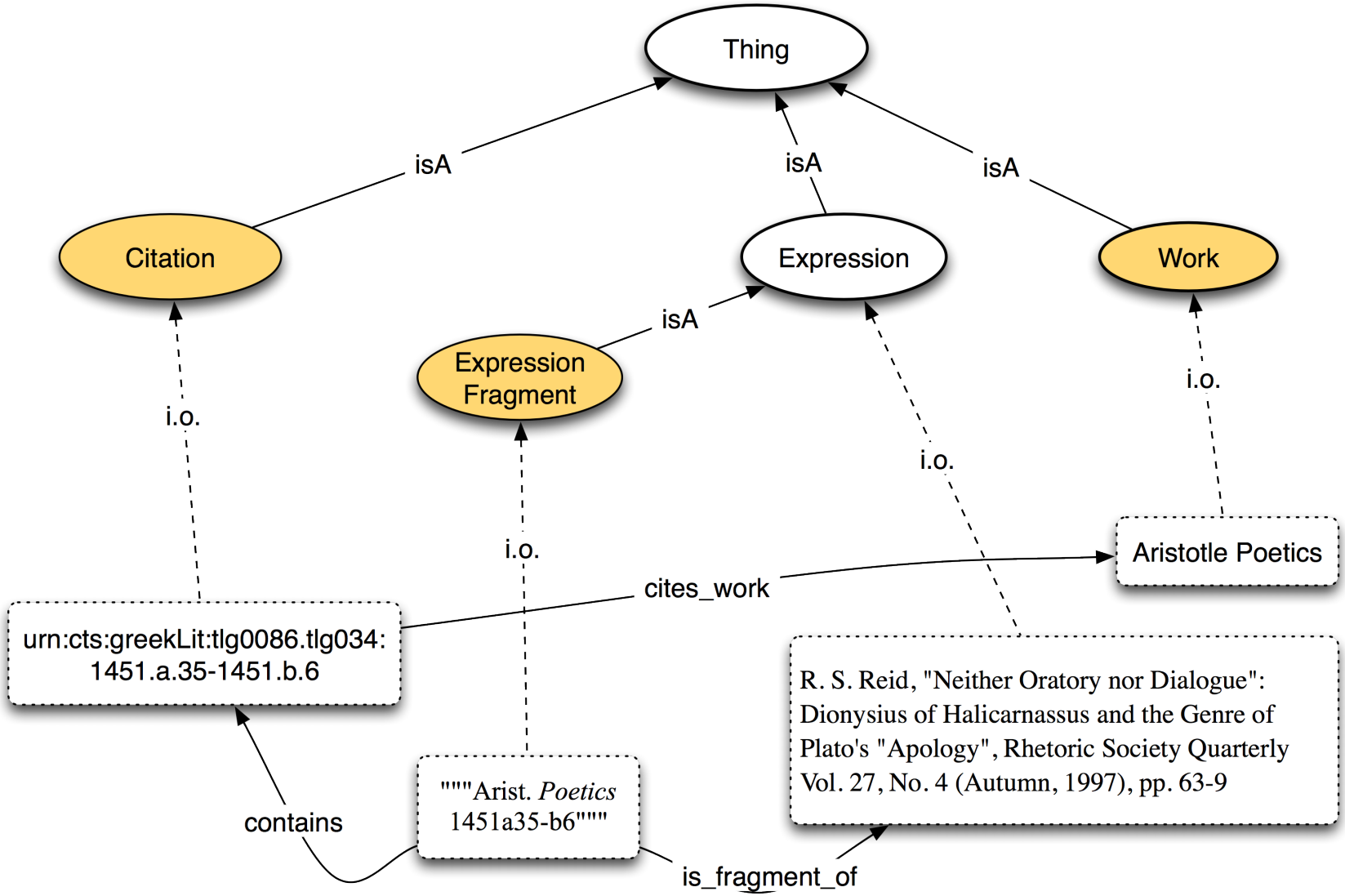
Example

- **Work**: The abstract content of Homer's Iliad
- **Expression**: The text of Iliad in the Italian translation by the poet Vincenzo Monti (XIX century)
- **Manifestation**: A copy of V. Monti, Iliad, Milano : Mondadori, 1995 [isbn: 8804394846]

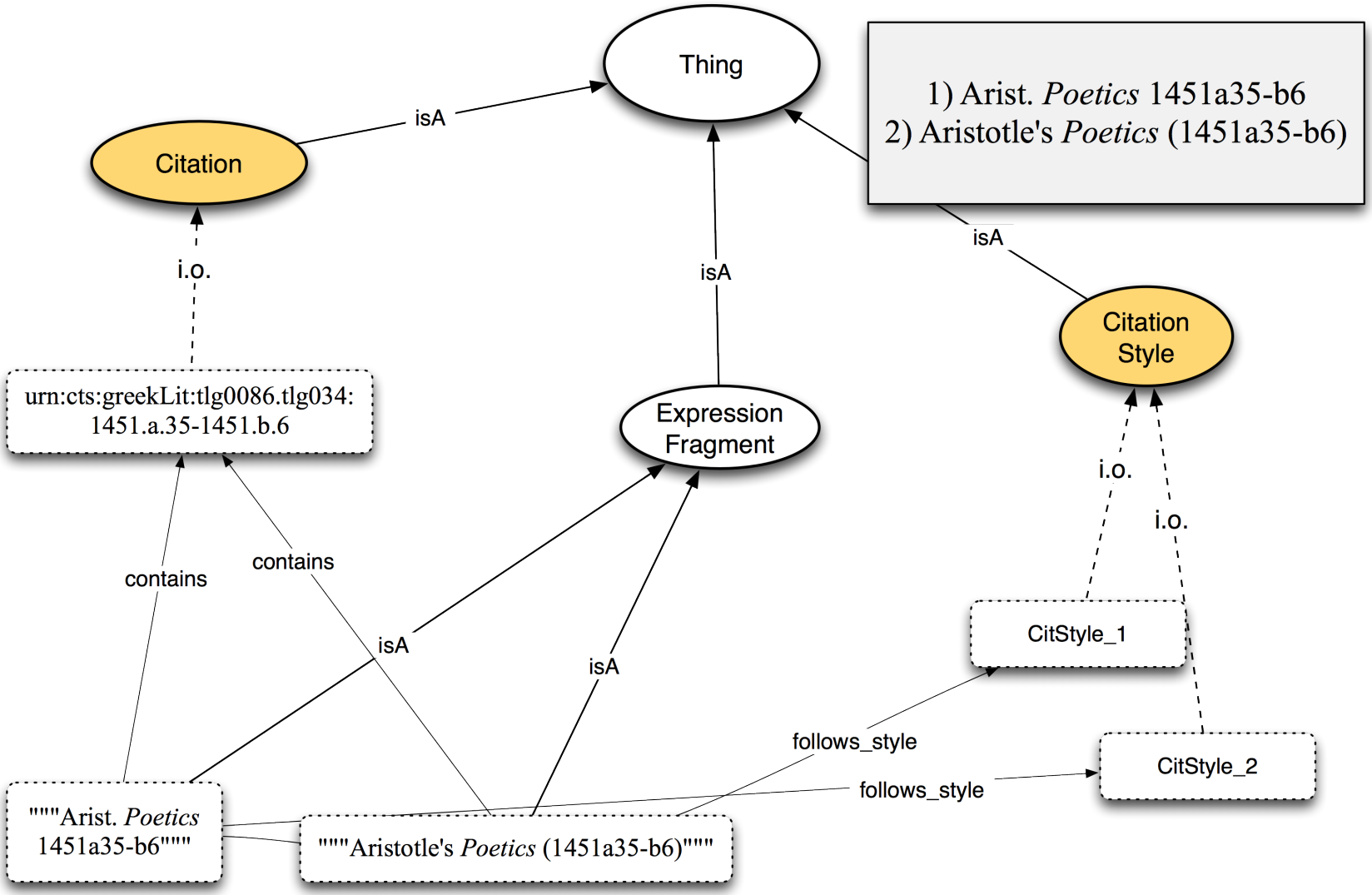
HuCit's Overview



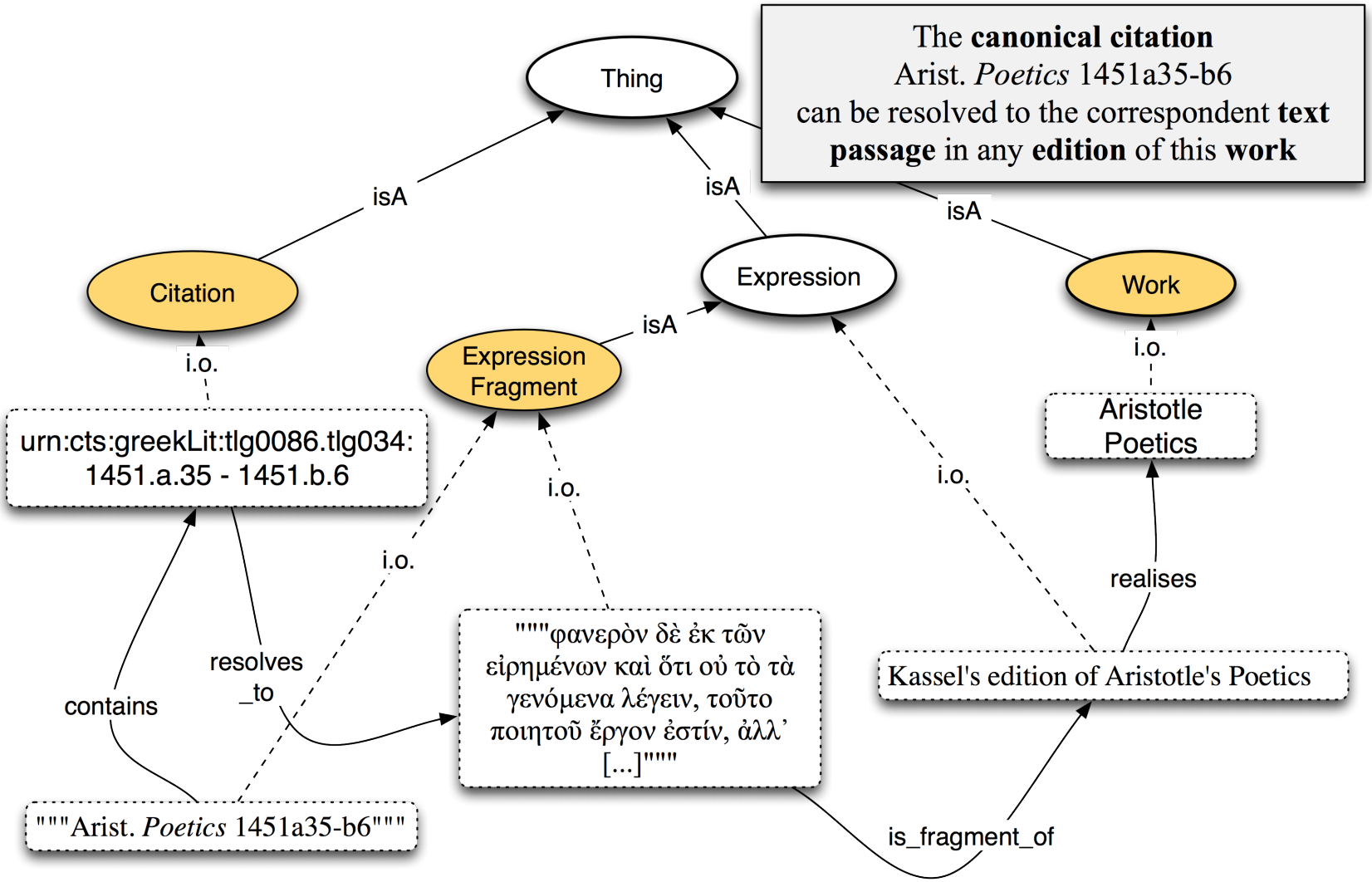
A Canonical Citation (macro level)



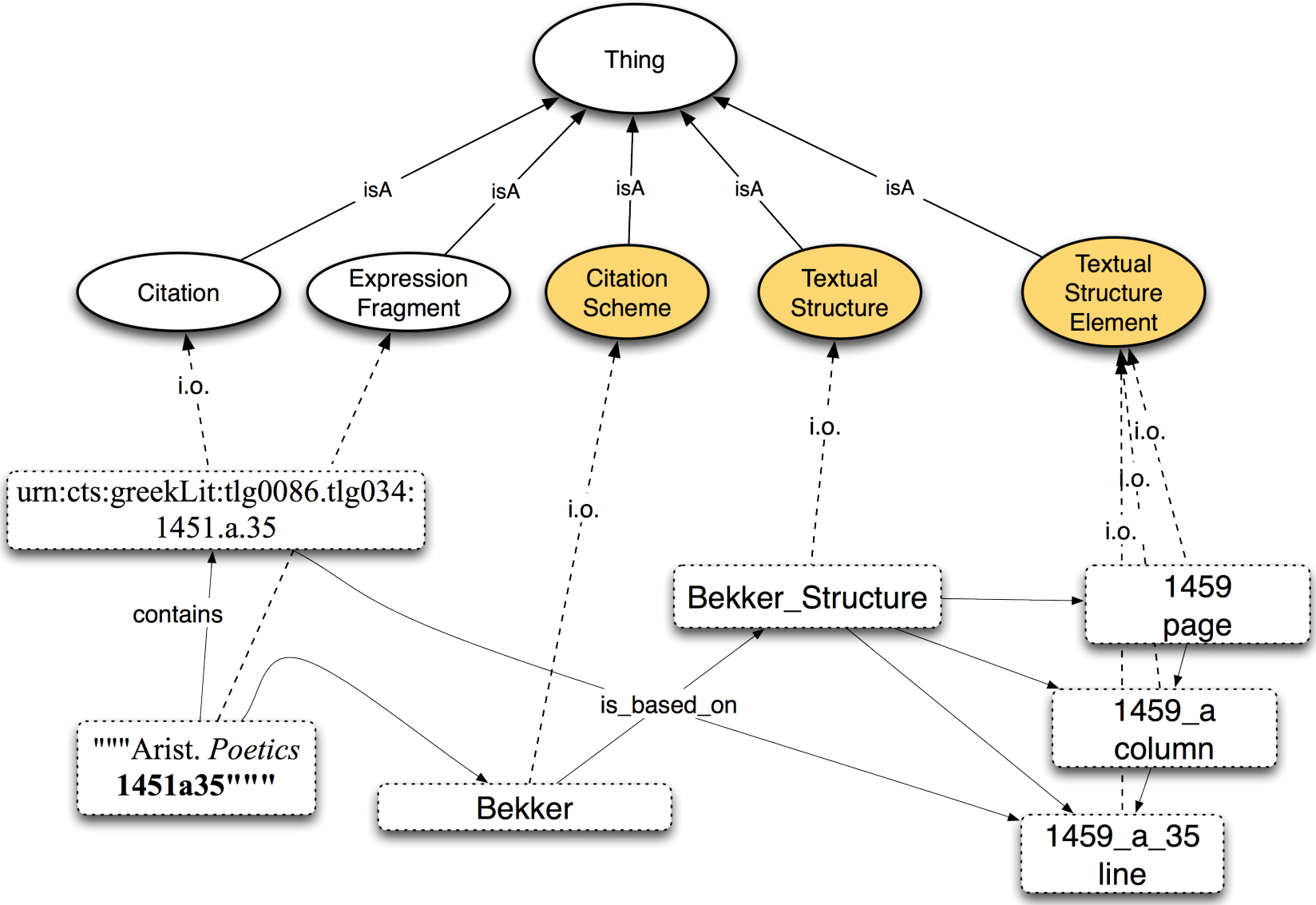
Citation Styles



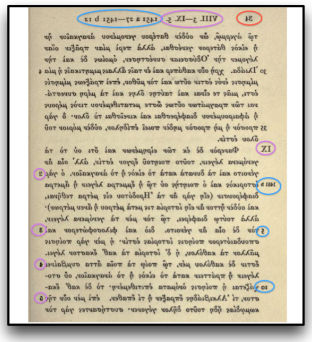
Canonical Citation as resolvable pointer



Citation Scheme and Textual Structure



Textual Structure and Expressions



Wrap up

Contribution

- a conceptual model of how canonical citations work
- a more abstract / application-independent model to encode semantics of canon. citations
- *interoperability* with other standards/protocols for canonical citations (CTS, CWKB)
- small ontology but mapped to other ontologies (CIDOC-CRM and FRBRoo)

Ontology population

- citations extracted from texts (JSTOR, APh)
- textual structures to be extracted from TEI encoded texts (Perseus)
- reasoning over the KB to support automatic extraction