

The banner features a row of six icons: a globe, a book, a network of nodes, a money bag with a Euro symbol, a scale of justice, and a bicycle. Below the icons, the text 'AIUCD 2021' is prominently displayed. Underneath, it reads 'DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale' and '10° congresso annuale PISA 19-22 gennaio'. On the right side, a list of topics is shown in various colors: 'DIGITAL PUBLIC HUMANITIES' (red), 'OPEN CULTURE' (orange), 'RETI SOCIALI' (yellow), 'TECH ECONOMY' (green), 'E-PARTICIPATION' (blue), and 'TECNOLOGIE ASSISTIVE' (purple). The background includes binary code and a classical building facade.

AIUCD 2021

DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale

10° congresso annuale **PISA** 19-22 gennaio

DIGITAL PUBLIC HUMANITIES
OPEN CULTURE
RETI SOCIALI
TECH ECONOMY
E-PARTICIPATION
TECNOLOGIE ASSISTIVE

Versione PROVVISORIA del contributo presentato al Convegno Annuale

DISCLAIMER

Questa versione dell'abstract non è da considerarsi definitiva e viene pubblicata esclusivamente per facilitare la partecipazione del pubblico al convegno AIUCD 2021

Il Book of Abstract contenente le versioni definitive e dotato di ISBN sarà disponibile liberamente a partire dal 19 gennaio sul sito del convegno sotto licenza creative commons.

Per un catalogo annotato della letteratura greca antica

Monica Berti

Universität Leipzig, Germania – monica.berti(«»)uni-leipzig.de

ABSTRACT

Questa proposta di poster illustra un progetto di estrazione, analisi e annotazione semi-automatica delle forme linguistiche pertinenti alle citazioni bibliografiche nelle fonti greche antiche. Scopo del progetto è la realizzazione di un catalogo annotato della letteratura greca basato sul linguaggio dei testi antichi e sullo sviluppo di una metodologia aperta e condivisa per la conservazione del patrimonio testuale storico. Il progetto offre dati aperti e accessibili che possono essere utilizzati e arricchiti in ambienti collaborativi per la ricerca e la didattica.

PAROLE CHIAVE

Annotazione Linguistica, CITE Architecture, Digital Libraries, Greco Antico, Named Entities, NLP

1. ANNOTAZIONI LINGUISTICHE E CITAZIONI ANTICHE

Cataloghi e indici cartacei degli autori e delle opere della letteratura greca antica sono il risultato di una lunga tradizione filologica ed editoriale che ha permesso – e ancora permette – di raccogliere informazioni preziose sul patrimonio testuale dell'antichità classica. Il limite di queste risorse è rappresentato dal fatto che esse generalmente comprendono soltanto forme lemmatizzate in latino o in lingue moderne, richiedendo dunque una lettura separata e analitica dei passi da esse raccolti al fine di comprendere la varietà e la complessità del linguaggio utilizzato dagli autori antichi per esprimere le citazioni bibliografiche (come, per esempio, l'onomastica degli autori e le descrizioni e i titoli delle opere). Anche se in ambiente digitale le annotazioni morfo-sintattiche del greco e del latino stanno crescendo [4, 3], mancano ancora collezioni complete e annotazioni arricchite delle cosiddette «named entities» (NEs) riguardanti le citazioni bibliografiche antiche di autori e opere estratte dalle fonti testuali [2].

Θεόπομπος [Theopompos] (lemma)

Lemma: Θεόπομπος [Theopompos]
lemma in [LOGEION](#)

Inflected form: Θεόπομπον [Theopompon] - [search in Kaibel Q](#)
inflected form in [TLG](#)
Named Entity Class: PER

10 . 8 [και αὐτὸς ἦν πολυφάγος προειρήκαμεν, παραθέμενοι Θεόπομπον ιστοροῦντα ἐν τῇ πέμπτῃ καὶ τριακοστῇ]
11 . 63 [[Θυατειρηνὸς](#) τὸ χωρὶς ὠτίων ποτήριον ὀνομακῆναι Θεόπομπον ἐν [Μήδῳ Φιλήμων Φάσματι](#): ἐπεὶ δ']

Inflected form: Θεόπομπος [Theopompos] - [search in Kaibel Q](#)
inflected form in [TLG](#)
Named Entity Class: PER

1 . 42 [γυναικῶν ἀνακείμενοι ὑπὸ τῷ αὐτῷ ἱματίῳ. Θεόπομπος, ἐπίνομεν μετὰ ταῦτα κατακείμενοι μαλακώτατ']
1 . 47 [ταῖς ἔξεισι τῶν πινόντων πλεῖστον χρόνον. Θεόπομπος δέ φησι παρὰ [Χίους](#) πρώτοις γενέσθαι]
1 . 61 [ὅτι ὁ Θεόπομπος ὁ [Χίος](#) τὴν ἄμπελον ἱστορεῖ εὐρεθῆναι]
2 . 18 [ἐκδιδούσης εἰς αὐτὸν κρήνης τινὸς πικρᾶς. Θεόπομπος δέ φησι περὶ τὸν [Ἐριγῶνα](#) ποταμὸν]
2 . 23 [καθάπερ καὶ τὸν [Φρύγα Μίδα](#) φησὶ Θεόπομπος, ὅτε ἐλεῖν τὸν [Σιληνὸν](#) ὑπὸ μέθης]
2 . 35 [δ' αὐτομάτοισιν τὰ μμιαῖκυλ' ἐφύετο πολλὰ. Θεόπομπος: τρώγουσι μύρτα καὶ πέποννα μμιαῖκυλα. [Κράτης](#):]
2 . 62 [διὰ τοῦ φ ἀσφάραγον ὀνομάζει. καὶ Θεόπομπος: κάπειτ' ἰδὼν ἀσφάραγον ἐν θάμνῳ τινί.]
2 . 77 [καὶ παρ' [Αἰσχύλῳ](#) διαβρέχεις τάρτυματα. καὶ Θεόπομπος δέ φησι: πολλοὶ μὲν ἀρτυμάτων μέδιμοι,]

Figura 1. Digital Athenaeus: Named Entities Concordance

Figura 1 mostra l'esempio della concordanza dei *Deipnosofisti* di Ateneo di Naucrati, la quale è liberamente consultabile online e offre diversi livelli di annotazione delle NEs (colori rosso e blu) pertinenti anche alle citazioni di autori e opere conservate nel testo greco (ed. Kaibel) per un totale di circa 23.000 token estratti, lemmatizzati e annotati. Questa risorsa si accompagna ad altri strumenti del progetto *Digital Athenaeus* (<http://www.digitalathenaeus.org>), i quali permettono una lettura ragionata del testo dei *Deipnosofisti* in quanto miniera preziosa di riferimenti bibliografici ad autori e opere della letteratura greca che risultano oggi per la maggior parte irrimediabilmente perdute [1]. Queste risorse possono

essere utilizzate non soltanto per la ricerca filologica e storiografica, ma anche per la didattica delle lingue classiche e dell'informatica umanistica ad esse applicata.

2. CATALOGAZIONE APERTA E CONDIVISA DELLA LETTERATURA GRECA ANTICA

Obiettivo del poster è la presentazione del lavoro di estrazione e annotazione semi-automatica di NEs al fine di creare un catalogo annotato della letteratura greca antica. Il testo di riferimento è rappresentato dai *Deipnosofisti* di Ateneo, ma lo scopo finale comprende l'elaborazione di un modello di annotazione e la produzione di dati che saranno utilizzati per l'annotazione di altri testi della letteratura greca, a partire dal lessico *Suda* e dal lessico di Arpocrazione. Mediante la presentazione di un esempio concreto, le fasi illustrate nel poster sono le seguenti: 1) estrazione e lemmatizzazione delle NEs; 2) annotazione delle singole NEs in classi generiche sulla base dell'interrogazione di «authority lists» esterne; 3) disambiguazione delle NEs pertinenti ad autori e opere antiche con sottoclassi semantiche e creazione di citazioni canoniche previste dalla CITE Architecture [1, 2]; 4) utilizzo della piattaforma di annotazione INCEpTION (<https://inception-project.github.io/>) per «relation extraction», «NE linking» e «co-reference resolution»; 5) esportazione dei dati in formati standard per la condivisione e il riuso (CSV, TEI XML, CTS e CITE urns, PAULA XML, CoNLL, TCF, WebAnno TSV 3, XMI). Figura 2 mostra un esempio di diversi livelli di annotazione testuale all'interno di INCEpTION [2].

The screenshot shows the INCEpTION web interface. At the top, it displays the document path 'berti: Deipnosophists/03-023.tsv' and '1-5 / 16 sentences [doc 171 / 250]'. Below this is a toolbar with navigation and editing icons. The main text area contains five lines of Greek text with various entities highlighted in green boxes. Each box contains a URN and the entity name in Greek. For example, the first line has 'urn:cts:greekLit:tg0510' for 'Τηλεκλειδης' and 'urn:cts:greekLit:tg0510.ath001' for 'Αμφικτύου'. The right-hand panel shows the 'Layer' set to 'Ancient Greek Catalog' and an 'Annotation' section with a 'Delete' button and a 'Clear' button. Below this, the 'Text' field contains 'Θεόπομπος' and the 'value' field contains 'urn:cts:greekLit:tg0513'.

Figura 2. Livelli di annotazione linguistica in INCEpTION (Ath. *Deipn.* 3.23)

Il progetto si inserisce nelle linee guida del tema «Open Culture» proposto dagli organizzatori di AIUCD2021 per le seguenti ragioni: 1) generazione di un corpus testuale annotato per il greco antico liberamente accessibile e riutilizzabile; 2) produzione di dati e risorse linguistiche (forme flesse, lemmi e NEs) concernenti nomi propri in greco antico che ancora mancano per la conservazione e la rappresentazione digitale del patrimonio testuale della civiltà classica; 3) utilizzo di una piattaforma open source sviluppata per la linguistica e liberamente utilizzabile da studiosi e studenti per il lavoro collaborativo e la condivisione dei risultati; 4) creazione di una risorsa didattica e scientifica che permette l'analisi contestuale del greco antico, la creazione di dati, lo sviluppo di basi di conoscenza («knowledge bases») e l'apprendimento delle metodologie di annotazione ed estrazione delle lingue storiche; 5) la promozione e l'implementazione di metodologie, dati, risorse linguistiche e ambienti collaborativi che offrono giuste opportunità nel mondo della ricerca, della formazione e della condivisione del patrimonio culturale.

3. RINGRAZIAMENTI

Si ringrazia la Deutsche Forschungsgemeinschaft (progetto nr. 434173983) per il sostegno finanziario al progetto.

BIBLIOGRAFIA

- [1] Berti, Monica et al. «Documenting Homeric Text-Reuse in the *Deipnosophistae* of Athenaeus of Naucratis». In Gabriel Bodard, Yanne Broux, e Ségolène Tarte (eds.). *Digital Approaches and the Ancient World*. BICS Themed Issue 59:2 (2016): 121-139. DOI: 10.1111/j.2041-5370.2016.12042.x
- [2] Berti, Monica. «Named Entity Annotation for Ancient Greek with INCEpTION». In *Proceedings of CLARIN Annual Conference 2019*. CLARIN, 2019: 1-4. https://office.clarin.eu/v/CE-2019-1512_CLARIN2019_ConferenceProceedings.pdf
- [3] Celano, Giuseppe G.A. «The Dependency Treebanks for Ancient Greek and Latin». In Monica Berti (ed.). *Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution*. De Gruyter, 2019: 279-297. DOI: 10.1515/9783110599572-016
- [4] Vatri, Alessandro, e Barbara McGillivray. «The Diorisis Ancient Greek Corpus». *Research Data Journal for the Humanities and Social Sciences* 3:1 (2018). DOI: 10.1163/24523666-01000013