

# La semantica della pagina scritta

Federico Boschetti

CNR-ILC

[federico.boschetti@ilc.cnr.it](mailto:federico.boschetti@ilc.cnr.it)

VENEZIA, 26 OTTOBRE 2023

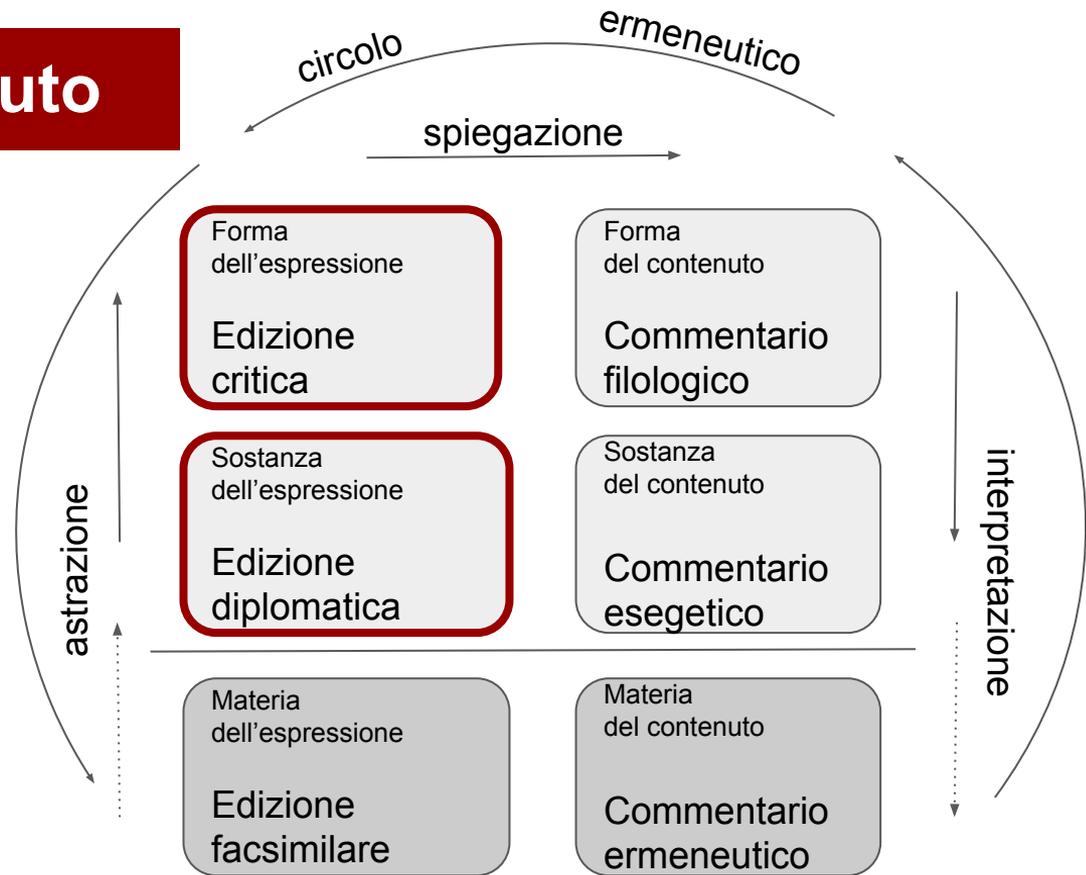
# **INTRODUZIONE**

# Un ponte fra beni culturali materiali e immateriali

L'Handwritten Text Recognition (HTR) permette di passare dalla rappresentazione digitale di beni materiali come manoscritti e altri text-bearing object alla rappresentazione digitale di beni immateriali come i testi

# **LA SEMIOTICA DELLA PAGINA SCRITTA**

# Espressione e contenuto

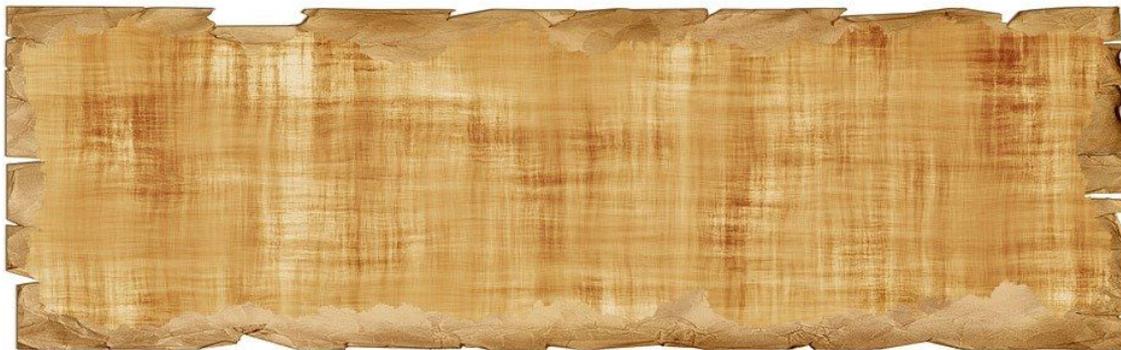


# La materia dell'espressione

La materia dell'espressione appartiene a una dimensione pre-linguistica e costituisce il **campo delle possibilità** su cui il segno linguistico si manifesta

# La materia dell'espressione

Studio del supporto scrittorio



Studio degli inchiostri

**INDICI (INDIZI), NON SIMBOLI**

Studio delle legature



# La sostanza dell'espressione

La sostanza dell'espressione è il luogo dove la **forma scritta** modella la **materia scrittoria** e dove tale materia resiste (confonde, oscura) la forma. Tale sostanza è l'oggetto principale dell'**edizione diplomatica**, dove vengono rappresentati gli **allografi** (piano -etico) e non i grafemi (piano -emico, dominio dell'edizione critica), le lacune senza ricostruzioni congetturali, le lettere puntate per indicarne l'incerta lettura

Gli **allografi** non hanno valore distintivo, ma sono validi indizi (=indici) di fenomeni culturali (tipo di scrittura) o personali (mano) di grande rilievo per il filologo

# La sostanza dell'espressione

CVI DONO LEPIDVM NOVVM LIBELLVM  
ARIDA MODO PVMICE EXPOLITVM?  
CORNELI, TIBI: NAMQUE TV SOLEBAS  
MEAS ESSE ALIQVID PVTARE NVGAS,  
5 IAM TVM, CVM AVSVS ES VNVS ITALORUM  
OMNE AEVVM TRIBVS EXPLICARE CARTIS,  
DOCTIS, IVPPITER, ET LABORIOSIS.  
QVARE HABE TIBI, QVICQVID HOC LIBELLI,  
QUALECVMQVE; QVOD, O PATRONA VIRGO,  
10 PLVS VNO MANEAT PERENNE SAECLO.

Cui dono lepidum nouum libellum  
Arida modo pumice expolitum?  
Corneli, tibi: namque tu solebas  
Meas esse aliquid putare nugas,  
5 lam tum, cum ausus es unus Italorum  
Omne aeuum tribus explicare cartis,  
Doctis, Iuppiter, et laboriosis.  
Quare habe tibi, quicquid hoc libelli,  
Qualecumque; quod, o patrona uirgo,  
10 Plus uno maneat perenne saeclo.

# La forma dell'espressione

La forma dell'espressione trova adeguata rappresentazione nell'**edizione critica**. L'edizione critica, infatti, è il luogo dove i segni hanno valore linguistico compiuto, a testo o in apparato. Fra tutte le possibilità di lettura lasciate aperte dalle edizioni diplomatiche, a prescindere dal significato, l'edizione critica restringe il campo a quelle letture che hanno maggior probabilità di dare senso. L'incertezza delle lettere puntate viene sciolta e le lacune vengono colmate con integrazioni congetturali. L'**edizione critica riguarda il testo come oggetto astratto**, cioè come oggetto liberato dalla materialità dei documenti che lo testimoniano



# La forma del contenuto

La forma del contenuto viene spiegata nel **commentario filologico** che accompagna l'edizione critica. Il commento filologico raccoglie il risultato di analisi morfosintattiche, semantiche, retoriche e stilistiche condotte sul testo e pertinenti a **giustificare le scelte** operate dall'editore per la *constitutio* della sua edizione. Nel commento filologico vengono riportati, quando richiesto, i *loci similes* (non necessariamente espressi in modo identico, ma vicini o identici per contenuto) a supporto di una lezione per chiarirne il senso e l'adeguatezza al contesto

# La forma del contenuto

VERG. ecl. 10, 69	▶ <b>Omnia uincit</b> Amor: et nos cedamus Amori."
VERG. app. ciris 437	▶ <b>Omnia uincit</b> amor: quid enim non uinceret ille?
OVID. Pont. 2, 7, 75	▶ <b>Omnia</b> deficiunt. animus tamen <b>omnia uincit</b> :
MANIL. astr. 4, 932	▶ Quas ratio, non pondus, habet: ratio <b>omnia uincit</b> .
AVIAN. fab. 1a, 4	▶ Femineus dulcis <b>omnia uincit</b> amor.
PROSP. epigr. 94, 8	▶ In quo persistens <b>omnia uincit</b> amor.
COLVMB. nav. 8	▶ Cuncta domat nisus, labor improbus <b>omnia uincit</b> .
ANTH. Lat. 931, 29	▶ Altus eram certe: Sortis furor <b>omnia uincit</b> ;

**VOCI CERCATE:**

- Homines • Amor • **Potentia amoris**

**COLLEGAMENTI TROVATI: 1–50 di 60**

Testo	Versi
anthologia Graeca, <i>liber 9</i> , 15	▶ <b>intero</b>
anthologia Graeca, <i>liber 9</i> , 420	▶ <b>intero</b>
anthologia Graeca, <i>liber 11</i> , 52	▶ 3–4
anthologia Graeca, <i>liber 16</i> , 103	▶ 6
anthologia Graeca, <i>liber 16</i> , 104	▶ <b>intero</b>
Catullus, <i>carmina</i> , 87	▶ <b>intero</b>
Propertius, <i>elegiae 1</i> , 1	▶ 3–4
Propertius, <i>elegiae 2</i> , 3	▶ 8
Propertius, <i>elegiae 2</i> , 8	▶ 40
Propertius, <i>elegiae 2</i> , 33	▶ 41–42
Propertius, <i>elegiae 2</i> , 34	▶ 24–26
Ovidius, <i>amores 2</i> , 18	▶ 11
Ovidius, <i>amores 3</i> , 11	▶ 34
epigrammata Bobiensia, 30	▶ 1–2
epigrammata Bobiensia, 36	▶ <b>intero</b>
aegritudo Perdicæ, 1	▶ 23–24 ▶ 33 ▶ 212 ▶ 270
Venantius Fortunatus, <i>carminum appendix</i> , 16	▶ 3–5
anthologia Latina, 9	▶ 13
anthologia Latina, 17 = Hosidii Getæ Medea	▶ 265
anthologia Latina, 83	▶ 96–98 ▶ 100
al-mu'allaqat, mu'allaqat 'imru'u l-qays	▶ 42–43
Bondie Dietaluti, <i>Gl'ochi col core stanno in tenzamento</i>	▶ 12–13
Cecco Angiolieri, <i>Sonetti</i> , Quanto un granel de panico è minore	▶ 12–14
Cecco Angiolieri, <i>Sonetti</i> , <i>Giuggiale di quaresima a l'uscita</i>	▶ 5–12
Chiario Davanzati, <i>Rime</i> , 55 (66)	▶ <b>intero</b>
Chiario Davanzati, <i>Rime</i> , 57 (68)	▶ 9–14

# La sostanza del contenuto

La sostanza del contenuto è resa esplicita dalla **traduzione scientifica** (intesa qui come *traduzione totale* nel senso di Torop) ed è oggetto del *commentario esegetico*. La traduzione scientifica costringe l'editore a esprimere non solo ciò che ha capito formalmente del testo, ma anche ciò che ha compreso, al punto di poterlo presentare, con le proprie parole, ad un pubblico completamente differente dal pubblico cui era destinato il messaggio originario

# La sostanza del contenuto

Bellandi et al.

Traduco e la Traduzione del Talmud

The screenshot displays the Traduco software interface, which is used for managing and translating Talmudic content. The interface is divided into several main sections:

- Resources:** A tree view on the left showing a hierarchy of documents, including chapters and specific sections like "19. La data del diluvio universale".
- Filter:** A central panel with search and filter options, including fields for Hebrew, Italian, Page, Paragraph, Note, Glossary type, and Glossary name.
- Table:** A central table with 11 rows, each representing a discussion item. The table columns include a row number, Hebrew text, Italian text, and a user ID. The highlighted row (row 7) contains the following text:

7	10b	7	הואש היתה נקודה עולה רשל ונשק	a Rosh haShanà <b>Sarà</b> , <b>Rachèl</b> e <b>Channà</b> furono ricordate da Dio e concepirono,	user-0033
---	-----	---	-------------------------------	---	-----------
- References:** A panel on the right showing a list of references for the selected item, including names like "Sarà", "Rachèl", and "Channà" with their respective user IDs.
- Translation suggestions:** A panel at the bottom right showing suggested translations for the selected item, including a rating system and a note about privacy.

(a)

(b)

(c)

# La materia del contenuto

La materia del contenuto è oggetto del **commentario ermeneutico**, che è volto a gettare luce sul testo in oggetto come prodotto di un contesto storico, antropologico, letterario e filosofico più ampio. La materia del contenuto è un *continuum* solo parzialmente sostanziato (e sostanziabile) in forma verbale. Il commento ermeneutico dovrebbe quindi tener conto dei dati o delle interpretazioni forniti da discipline come l'archeologia, la storia dell'arte, l'antropologia storica, che cercano di ricostruire il contesto extralinguistico e il clima culturale del mondo in cui i testi hanno avuto origine

# La materia del contenuto



# Collegare i livelli dei due piani

Il paradigma hjelmsleviano aiuta a tenere distinti i diversi livelli di astrazione sul piano dell'espressione e sul piano del contenuto e a stabilire il modo in cui collegare le unità degli uni rispetto alle unità degli altri

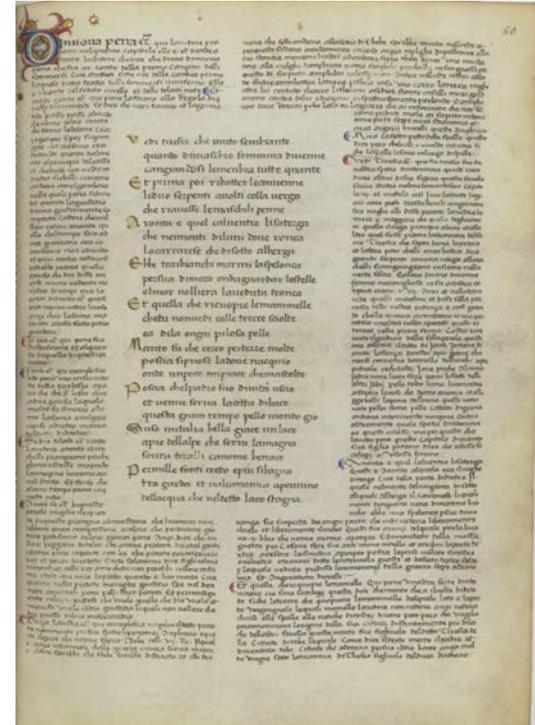
In presenza di facsimile, edizione diplomatica ed edizione critica, come nel progetto GreekSchools, i grafemi dell'edizione critica si mappano sui glifi dell'edizione diplomatica e questi ultimi si mappano su regioni d'interesse del facsimile, escludendo un collegamento diretto fra edizione critica e facsimile

Parimenti, in presenza di commentario filologico, esegetico ed ermeneutico il nome di una città antica (ad esempio) si mappa sulla sua descrizione storica, la quale si mappa a sua volta sulle moderne coordinate GIS della città moderna

# **LA SEMANTICA DELLA PAGINA SCRITTA**

# Testi e paratesti

La pagina può contenere più flussi testuali, spesso costituiti da un testo principale e da uno o più paratesti



# Layout complessi

I manoscritti, così come le edizioni a stampa, possono presentare layout complessi, ordinati o caotici e hanno due livelli di significazione:

- **significazione simbolica degli elementi verbali**, che hanno uno sviluppo **temporale e lineare**
- **significazione indicale, iconica o simbolica degli elementi non verbali del documento**, che hanno uno sviluppo **spaziale, bi- e/o tridimensionale**



# Layout complessi



- Numerazione
- Informazione di contesto
- Text-bearing object
- Testo di iscrizione latina
- Testo di iscrizione greca
- Traduzione dal greco al latino

# SegmOnto

## SegmOnto: A Controlled Vocabulary to Describe the Layout of Pages

How to cite

Simon Gabay, Ariane Pinche, Kelly Christensen, Jean-Baptiste Camps, Nicola Carboni, *SegmOnto, A Controlled Vocabulary to Describe the Layout of Pages*, version 0.9, Genève/Lyon/Paris, 2023, <https://segmonto.github.io/>.



SegmOnto offers a controlled vocabulary to describe the content of books or manuscripts pages order to homogenise the data required by layout analysers. This project follows a double objective:

- Mutualise data to train stronger models on various layouts.
- Design a standardised pipeline for text extraction, from page scans to structured documents

SegmOnto is thought as a generalist description scheme, covering written documents produced since the apparition of the codex, but it has been designed using mainly western and middle eastern documents.

# SegmOnto



- NumberingZone:item
- MarginTextZone:metadata
- GraphicZone:figure
- MainZone:latin-block
- MainZone:greek-block
- MarginTextZone:translation

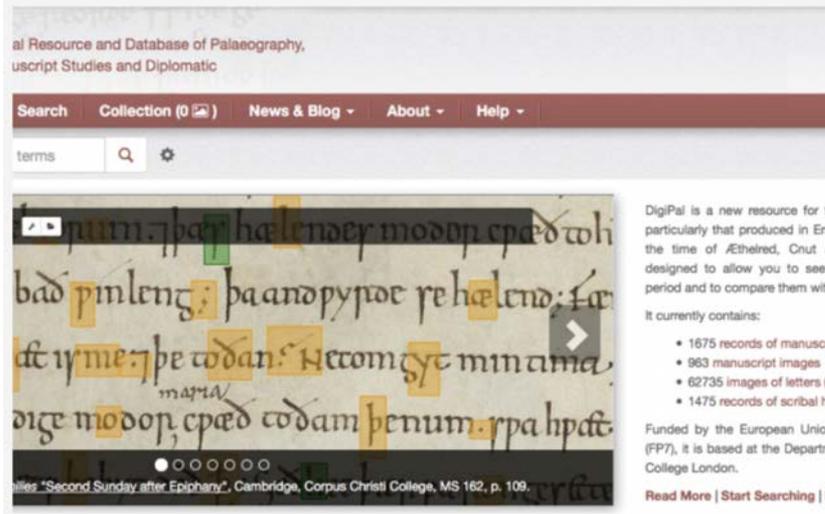
# Paleografia digitale

La paleografia (digitale) studia

- i tipi di scrittura
- le differenti mani
- gli allografi
- la probabilità dei segni alfabetici

La paleografia (digitale) è scienza della sostanza dell'espressione

## DigiPal



DigiPal is a new resource for the study of medieval handwriting, particularly that produced in England during the time of Æthelred, Cnut and Harold Godwinson. It is designed to allow you to see individual letters and to compare them with other examples from the same period and to compare them with other examples from the same period.

It currently contains:

- 1675 records of manuscripts
- 963 manuscript images
- 62735 images of letters
- 1475 records of scribal hands

Funded by the European Union (FP7), it is based at the Department of Digital Humanities, University of London.

[Read More](#) | [Start Searching](#) | [Help](#)

The Digital Resource and Database for Palaeography, Manuscript Studies and Diplomatic (DigiPal) aims to bring new methods in Digital Humanities to the study of medieval handwriting in its diplomatic and manuscript context by combining digital catalogues, descriptions of handwriting, and images of documents and their constituent letter-forms.

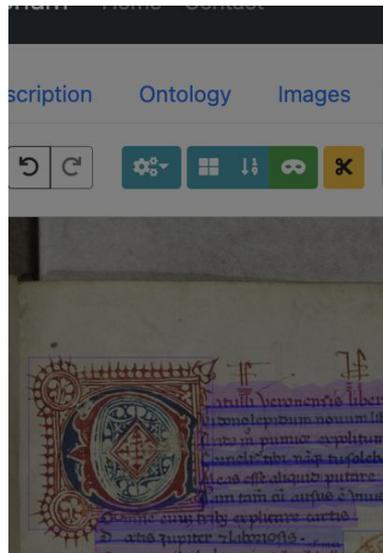
<https://eadh.org/projects/digipal>

# Lo scopo determina la discretizzazione

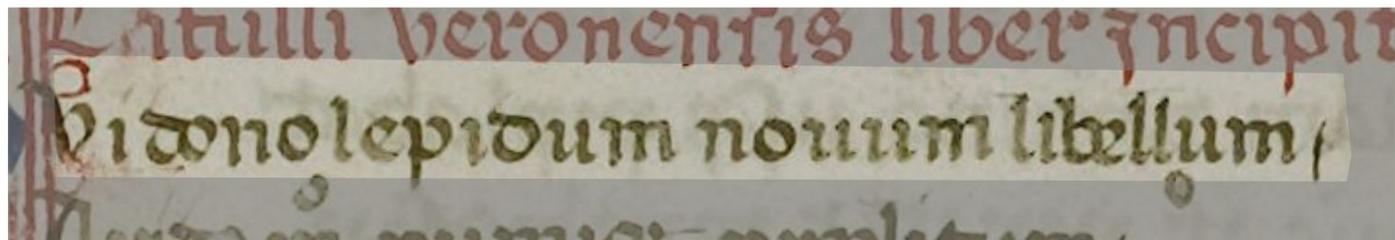
La digitalizzazione, così come la tipografia, è un processo di discretizzazione del continuo

Il modo in cui discretizzare il continuo non è intrinseco all'oggetto della digitalizzazione ma è determinato dalla **percezione** e dalla **rilevanza** dei fenomeni da **discriminare** rispetto all'**obiettivo di ricerca**

# Mappare il testo sul documento



← → 🗨 Line #2



vi dono lepidum nouum libellum,

by (eScriptorium) on Wed Sep 20 2023 15:15:39 GMT+0100

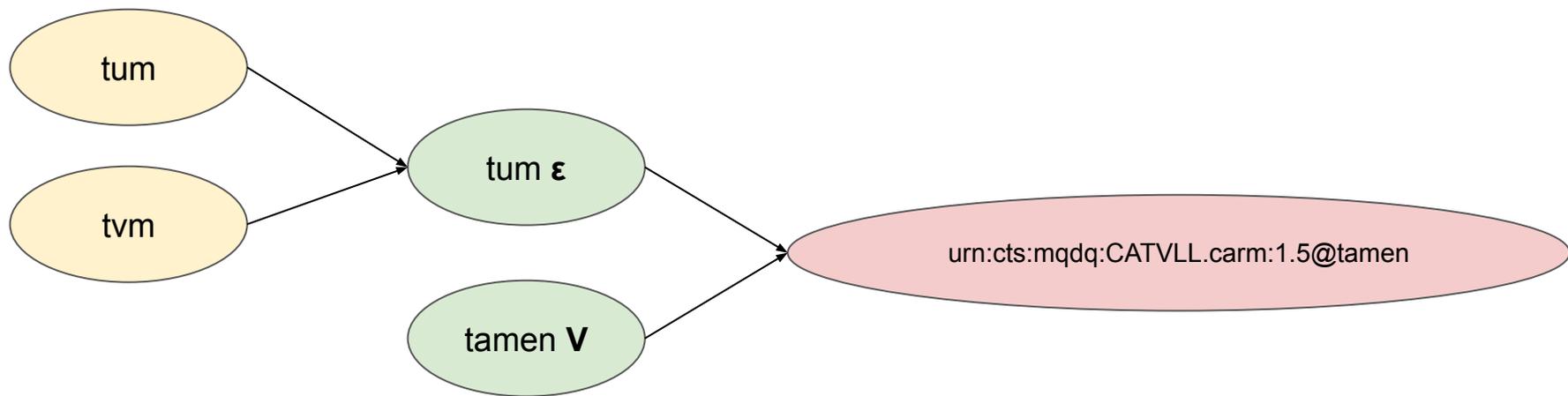
# Mappare il documento sul testo



Cui dono lepidum novum libellum

**Problemi di granularità**

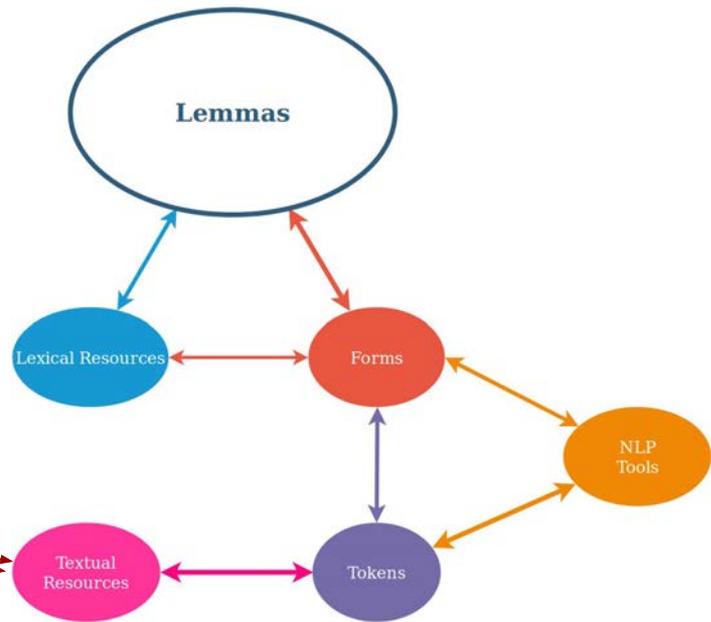
# Allografie e varianti



# Dal documento al testo, dal testo alla lingua



Cui dono epidum nouum libellum  
 Arida mode pumice expolitur?  
 Corneli, tibi: namque tu solebas  
 Meas esse aliquid putare nugas,  
 5 Iam **tum**, cum ausus es unus Italorum  
 Omne aeuum tribus explicare cartis,  
 Doctis, Iuppiter, et laboriosis.  
 Quare **habe tibi**, quicquid hoc libelli,  
 Quaecumque; quod, o patrona uirgo,  
 10 Plus uno maneat **perenne** saeclo.  
<https://mizar.unive.it/mqdg>



<https://lila-erc.eu>

# La digitalizzazione come processo infinito

La digitalizzazione, in quanto attività ermeneutica, è un processo infinito che va dalla pre-comprensione del tutto (standard, best practices) verso la comprensione delle parti (codifica, analisi linguistiche, NER: edizioni scientifiche digitali) e dalla comprensione delle parti (indici e concordanze) a una più ampia comprensione del tutto (statistiche, grafici: commentari filologici, esegetici, ermeneutici)

# La digitalizzazione come tradizione

- Il processo di digitalizzazione si inserisce nel flusso della **tradizione testuale**
- Siamo all'inizio della **quarta rivoluzione** (secondo G. Roncaglia: dall'oralità alla scrittura, dal rotolo al codice, dal manoscritto al libro a stampa e dalla stampa al digitale)
- Come in tutti i momenti di passaggio è molto forte il rischio di una **nuova vulgata** (P. Mastandrea), fortunatamente attenuato dalla digitalizzazione di testimoni multipli
- Non basta salvaguardare la **tradizione testuale** (piano dell'espressione): è necessario proteggere anche le **tradizioni di studio** (piano del contenuto)

**CONCLUSIONE**

# Conclusione

- Ogni edizione (digitale) è un'interpretazione di segni che stanno sulla pagina (e che si possono estendere al text-bearing object nella sua tridimensionalità di oggetto fisico in contesto)
- Siccome l'interpretazione è un processo infinito, un'edizione digitale deve essere **aperta**, **dinamica** e **tracciabile**, per poter evolvere in modo **libero** verso **direzioni impreviste** ma **verificabili**
- Siccome l'interpretazione riguarda la significazione nella sua completezza, l'edizione scientifica digitale, oltre ad avere una componente facsimilare, diplomatica e critica, dovrebbe essere accompagnata da un commentario filologico, esegetico ed ermeneutico