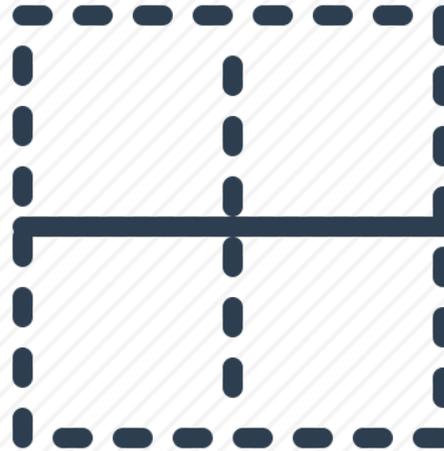


Paolo Monella

La codifica di livello zero



La codifica del testo. Due proposte tra teoria e prassi,

Università Cattolica

del Sacro Cuore, Brescia

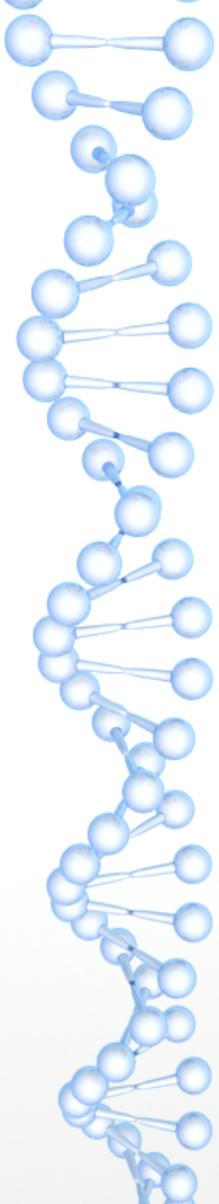
Facoltà di Scienze Linguistiche

e Letterature Straniere

17.05.2022



European Research Council
Established by the European Commission



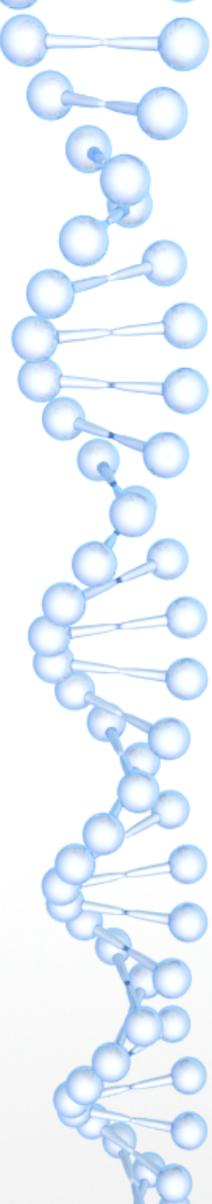
European Research Council

Established by the European Commission

ERC PAGES (AdG 2019 n° 882588)

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 882588)

Sommario



Sommario

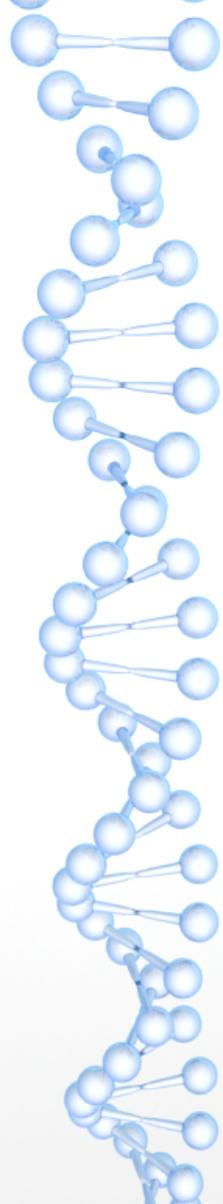
- La **trascrizione diplomatica** delle fonti primarie
- **Processamento**
 - La questione
 - Le attuali soluzioni
 - Processabilità, modellazione, documentazione
- La **tabella dei segni** (T. Orlandi)
 - Alcuni esempi di tabella dei segni



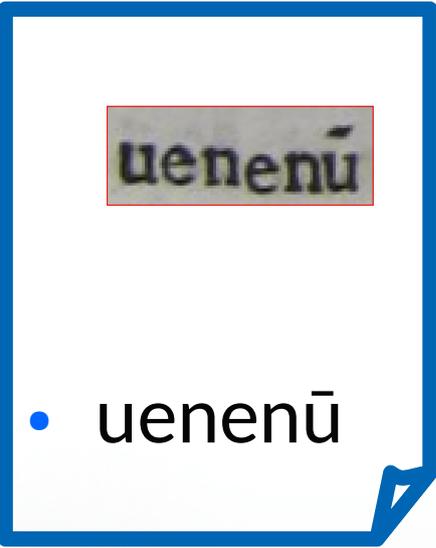
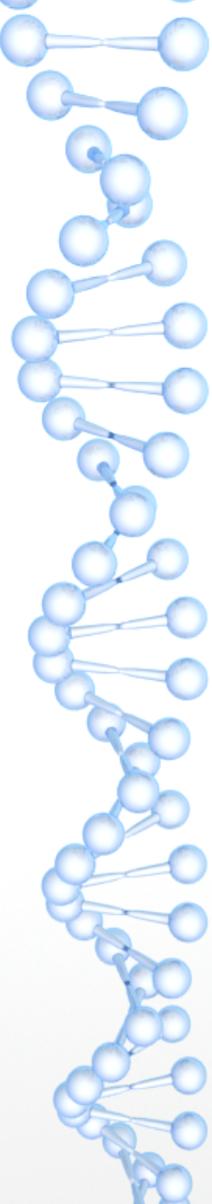
**La trascrizione diplomatica
delle fonti primarie**

Trascrizione diplomatica

uenenũ



Trascrizione diplomatica

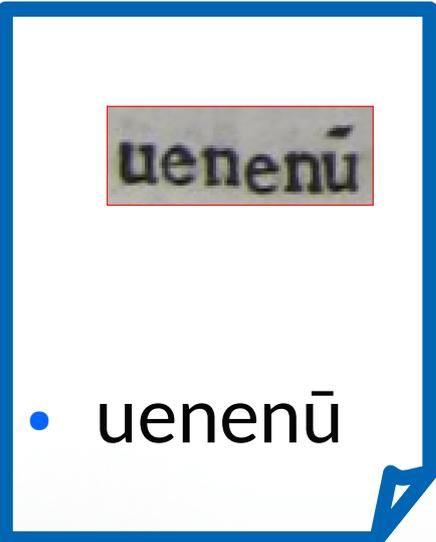
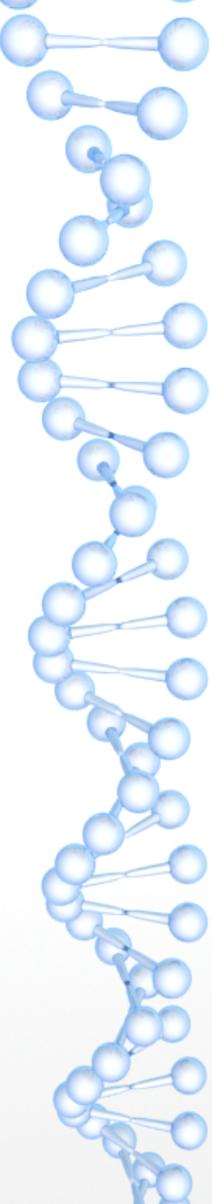


uenenū

- uenenū

Diplomatica

Trascrizione diplomatica



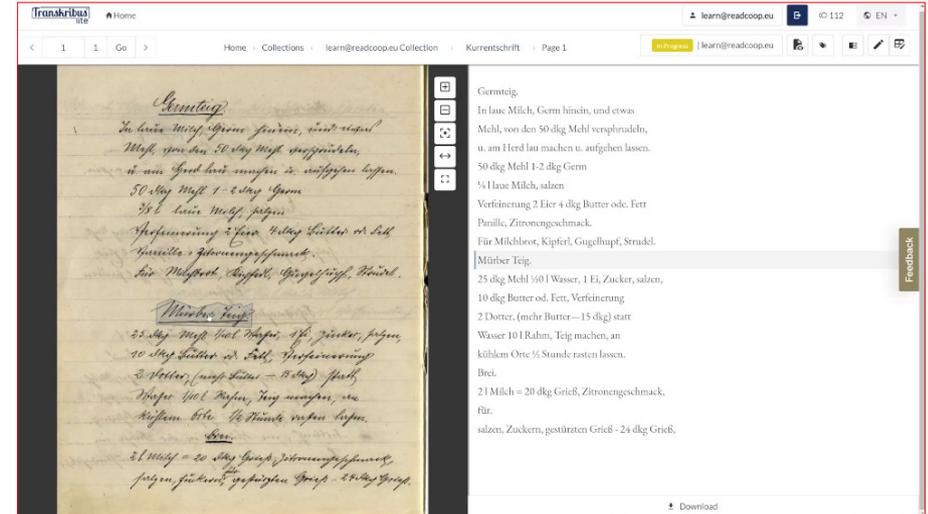
uenenū

- uenenū

Diplomatica

- Manuale o HTR

Trascrizione diplomatica



uenenū

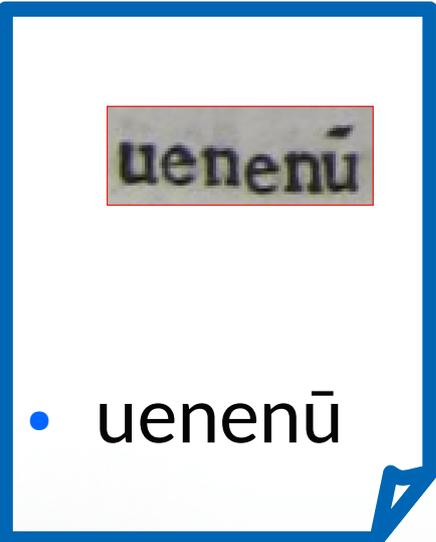
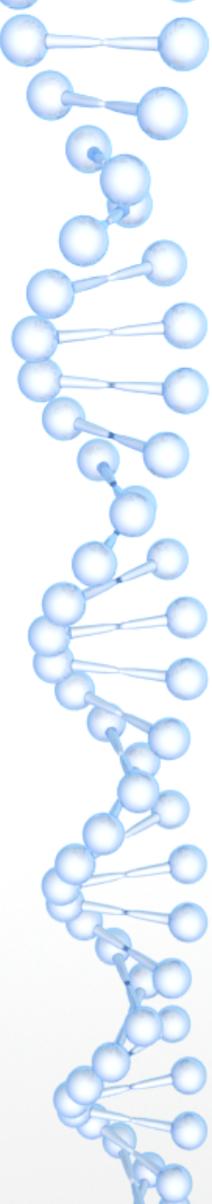
- uenenū

Diplomatica

- Manuale o **HTR**

Handwritten Text Recognition
(OCR per la scrittura a mano)

Trascrizione diplomatica



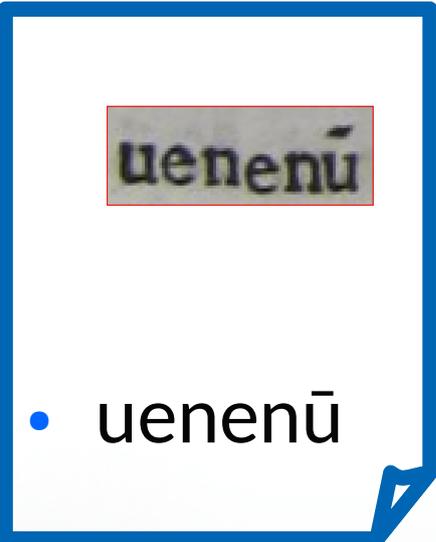
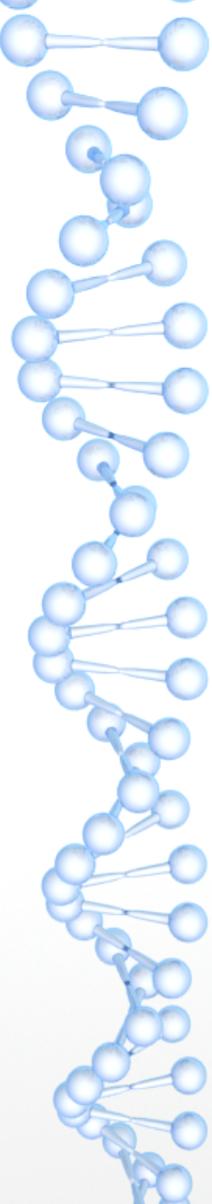
uenenū

- uenenū

Diplomatica

- Manuale o HTR
- Visualizzazione
- Processamento

Trascrizione diplomatica

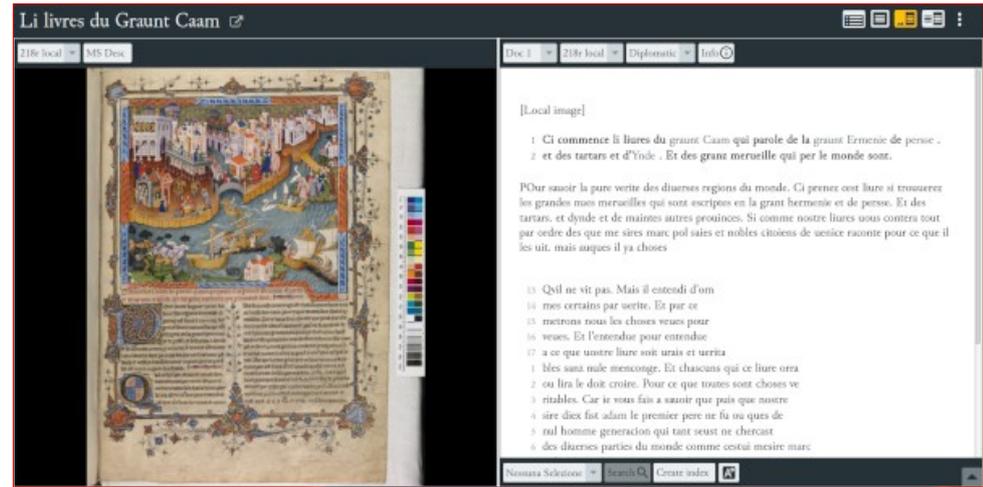


uēnenū

- uēnenū

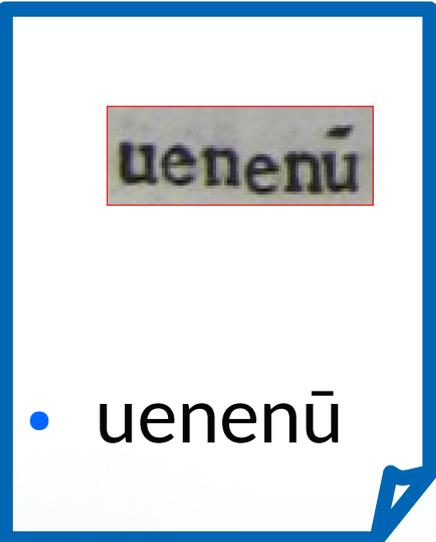
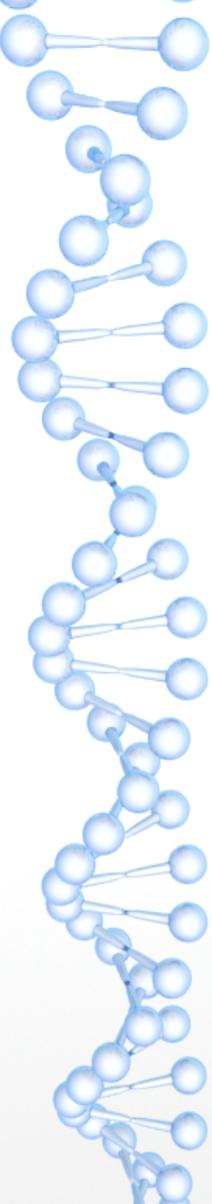
Diplomatica

- Manuale o HTR
- **Visualizzazione**
- Processamento



EVT
(Edition Visualization Technology)

Trascrizione diplomatica



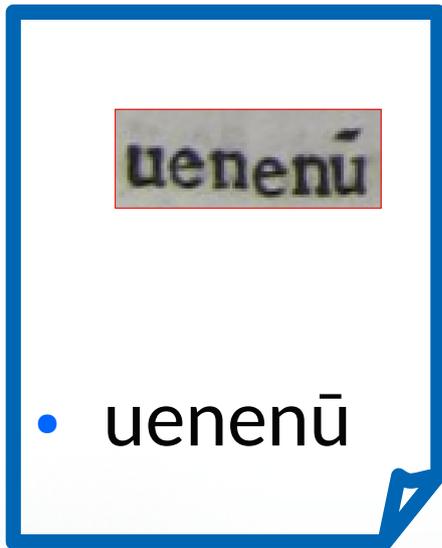
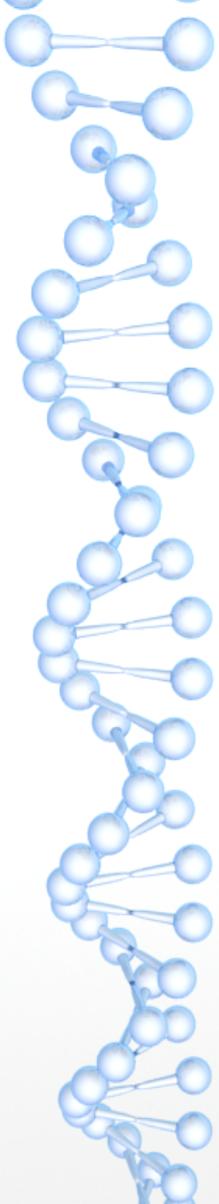
uenenū

- uenenū

Diplomatica

- Manuale o HTR
- Visualizzazione
- **Processamento**

Trascrizione diplomatica



Diplomatica

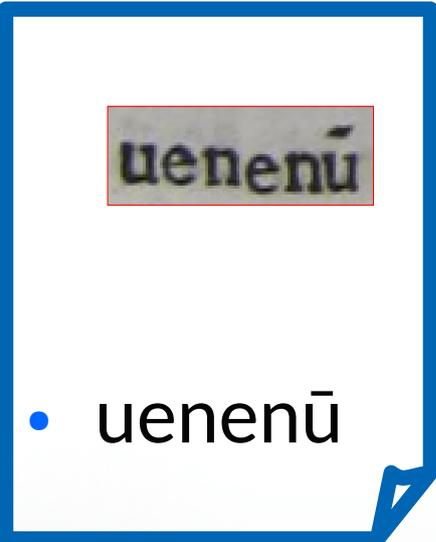
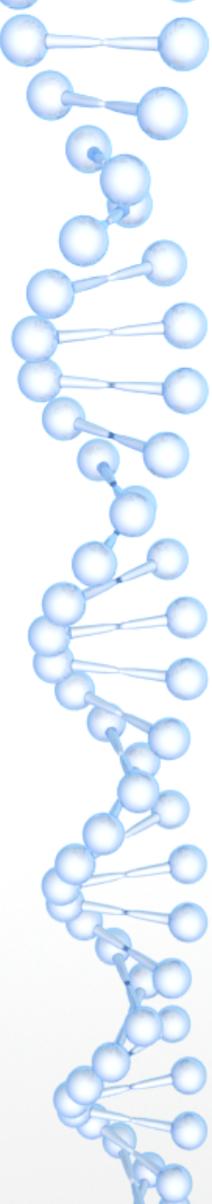
- Manuale o HTR
- Visualizzazione
- **Processamento**

Processabilità
(interoperabilità)



**Processamento:
la questione**

Processamento: la questione



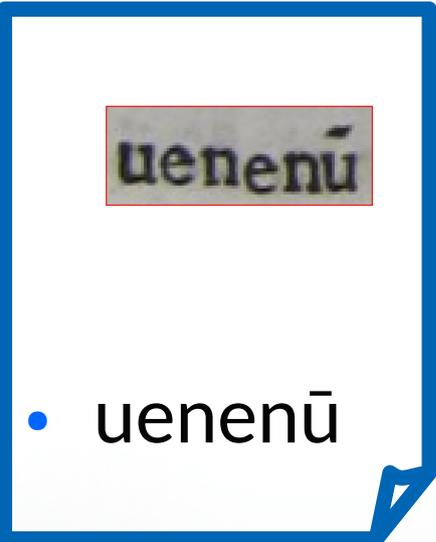
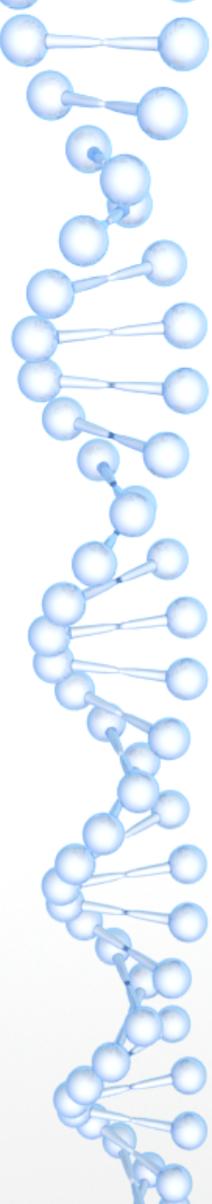
uenenū

- uenenū

Diplomatica

- Manuale o HTR
- Visualizzazione
- **Processamento**

Processamento: la questione

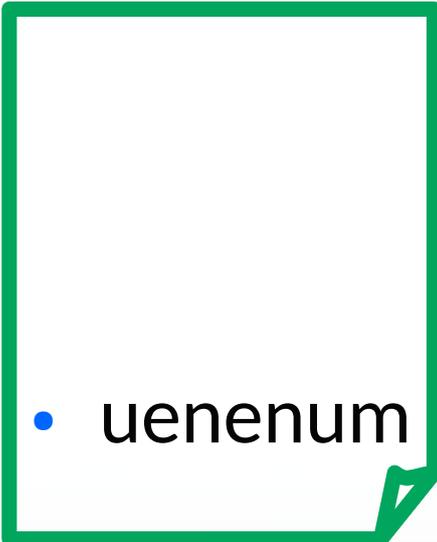


uenenū

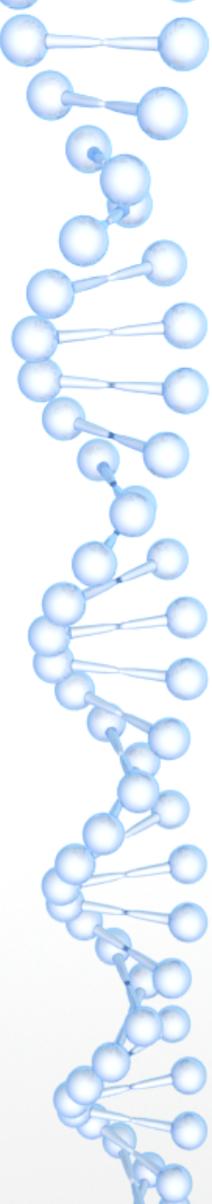
• uenenū

Diplomatica

- Manuale o HTR
- Visualizzazione
- **Processamento**

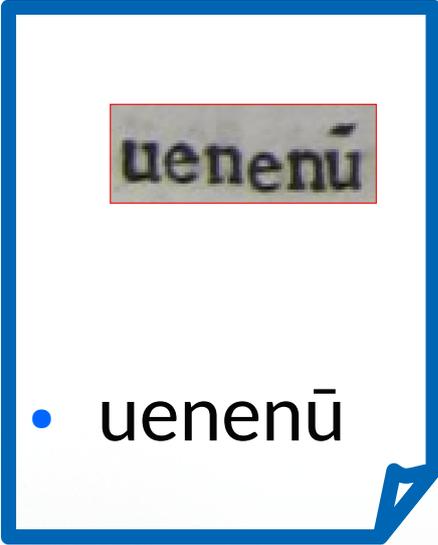


• uenenum



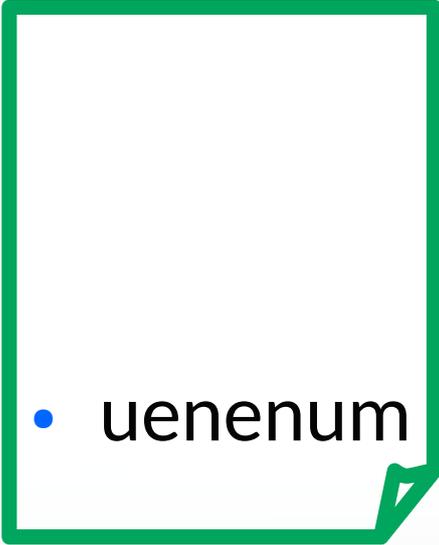
Processamento: la questione

- Processamento
 - Ricerca
 - Collazione
 - NLP (lemma, PoS etc.)
 - Statistiche (dist. reading)



uenenū

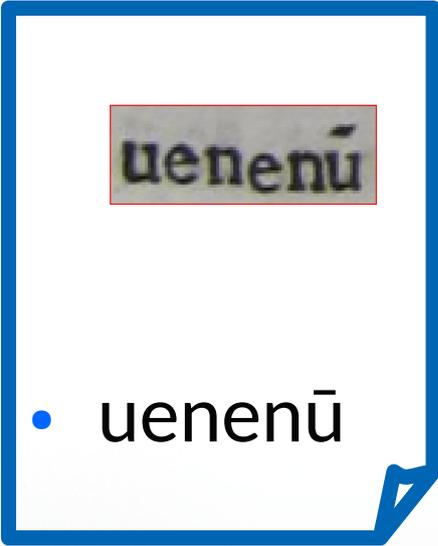
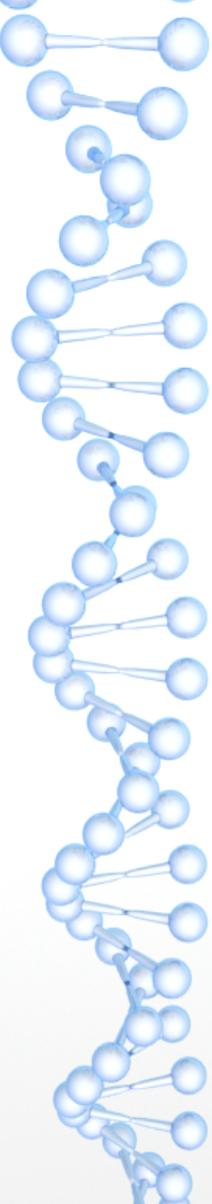
• uenenū



• uenenum

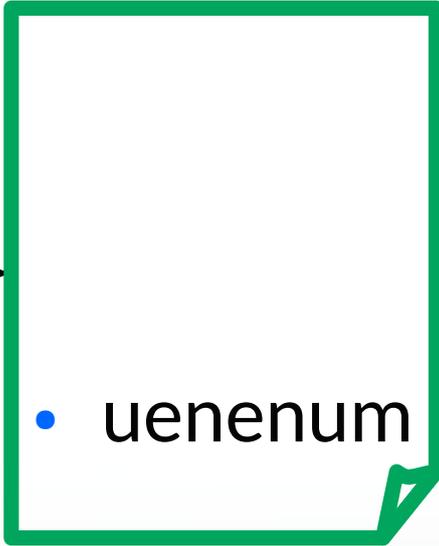
Processamento: la questione

- Processamento
 - **Ricerca**
 - Collazione
 - NLP (lemma, PoS etc.)
 - Statistiche (dist. reading)



uenerū

• uenerū



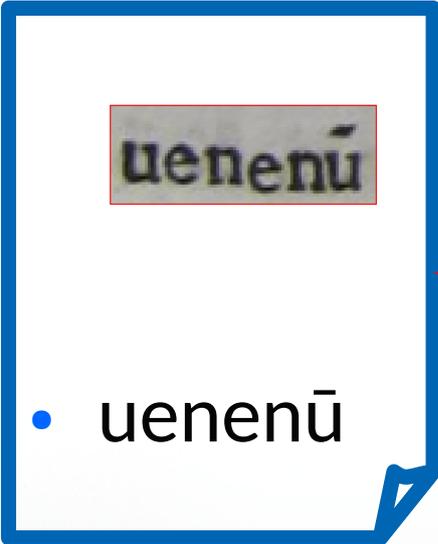
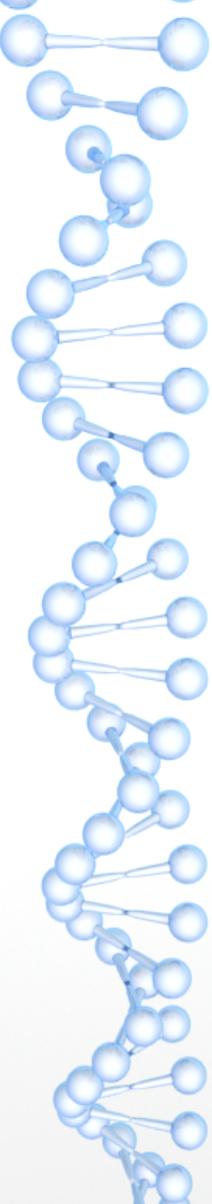
• uenerum



venenum

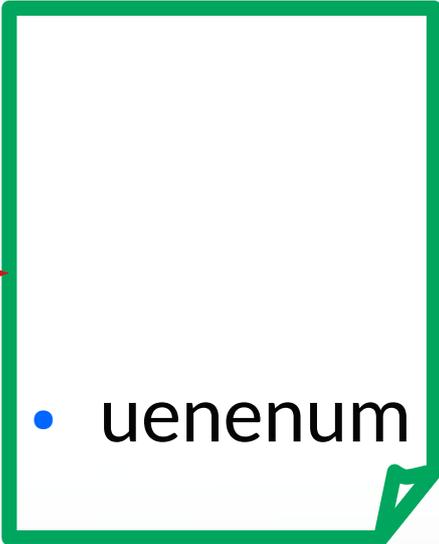
Processamento: la questione

- Processamento
 - Ricerca
 - **Collazione**
 - NLP (lemma, PoS etc.)
 - Statistiche (dist. reading)



uenenū

• uenenū

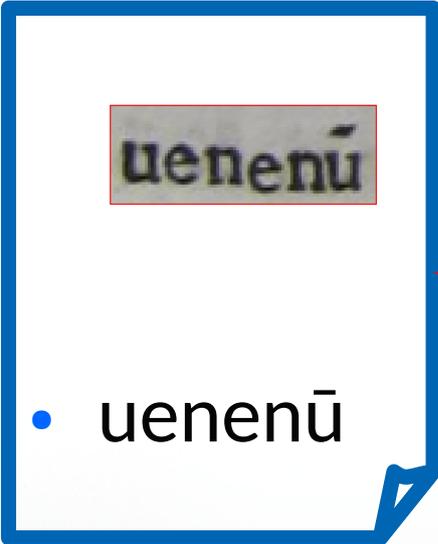
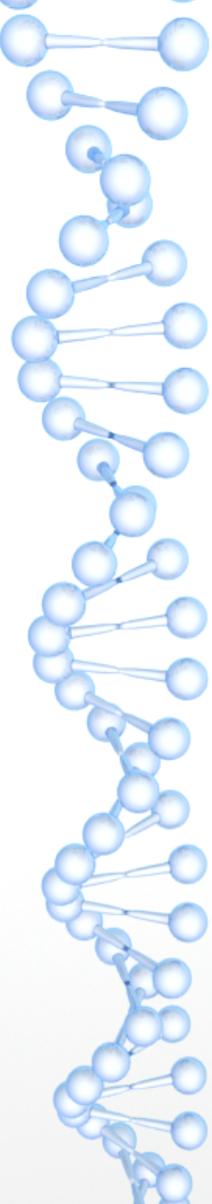


• uenum

112 excedit] excedit corr. ex
exceditis R n70
114 obicitur] obiceretur V S n71
114 sunt] sint S n72

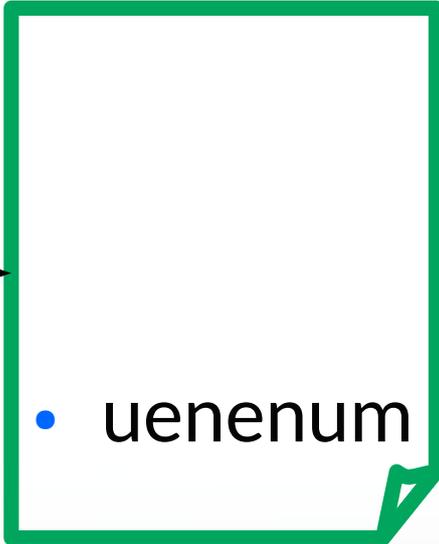
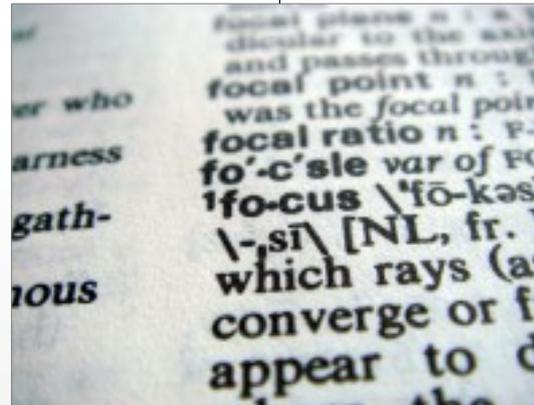
Processamento: la questione

- Processamento
 - Ricerca
 - Collazione
 - **NLP (lemma, PoS etc.)**
 - Statistiche (dist. reading)



uenenū

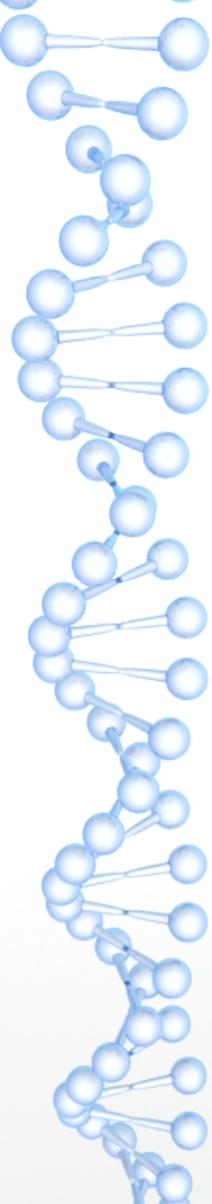
• uenenū



• uenum

Processamento: la questione

- Processamento
 - Ricerca
 - Collazione
 - NLP (lemma, PoS etc.)
 - **Statistiche (dist. reading)**

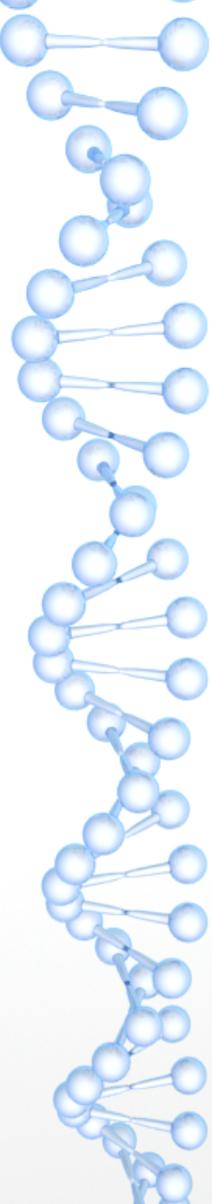


uēnēñ

• uēnēñ

• uēnēnum

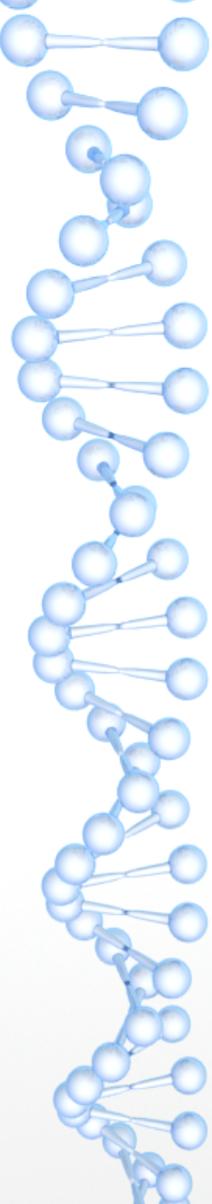




Processamento: la questione

- Mio interesse: **manoscritti** medievali europei
 - ...e prime stampe (che imitano i manoscritti)
- In linea di principio si applica ad **ogni sistema** grafico

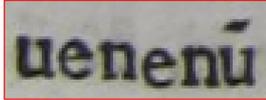
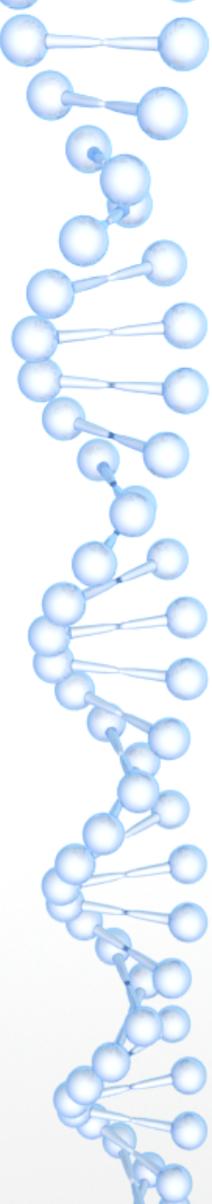
**Processamento:
le attuali soluzioni**



Unicode (la raccomandazione TEI)

- Soluzione per i testi digitali di **nuova** creazione
- O per **visualizzare** sistemi grafici pre-moderni
- Non basta per **processare** sistemi grafici pre-moderni
 - Allografi
 - **f** (U+017F) / **s** (U+0073; ASCII 115)
 - Ho 'detto' al computer che l'uno corrisponde all'altro (varianti del grafema <s>)?
 - Sistema grafematico
 - **u** (U+0075; ASCII 117)
 - Ho codificato se "copre" (o no) <u> e <v>?

Trascrizione 'normalizzata' manuale



uenenū

- uenenū

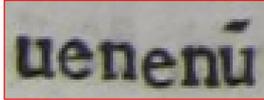
Diplomatica

- Visualizzazione
- Processamento

Trascrizione 'normalizzata' manuale

- venenum

Normalizzata



uenenū

- Processamento
 - Ricerca
 - Collazione
 - NLP (lemma, PoS etc.)
 - Statistiche (dist. reading)

- uenenū

Diplomatica

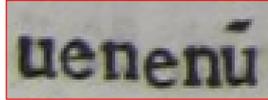
- Visualizzazione
- Processamento

Trascrizione 'normalizzata' manuale

• **v**enenum

Normalizzata

- Processamento
 - Ricerca
 - Collazione
 - NLP (lemma, PoS etc.)
 - Statistiche (dist. reading)



uenenū

• **u**enenū

Diplomatica

- Visualizzazione
- Processamento

Non generata
dal computer:
doppia trascriz.
manuale

Pre-processamento per specifica applicazione

- Collazione: **pre-processamento** per 'normalizzare' il testo
 - Oppure: configurare il software di collazione per **ignorare** *alcune* discrepanze (maiuscole, u/v etc.)
- Lo stesso per il software di ricerca, NLP, statistiche...
- Soluzioni '**fai da te**' e '**usa e getta**'
 - ...per un problema di fondo comune



**Processabilità,
modellazione,
documentazione**

Processabilità, modellazione, documentazione



WANTED

**DEAD OR ALIVE
REWARD
\$4,000**

- **Documentare** le pratiche di **modellazione** (trascrizione) e 'normalizzazione' del proprio progetto
 - Cosa sto trascrivendo? (Grafemi/allografi)? Come?
 - Cosa corrisponde a cosa (f/s)?
- Documentazione
 - In prosa ('specifiche' per futuri programmatori)
 - **Formale** (codice software, tabelle)

Processabilità, modellazione, documentazione



WANTED

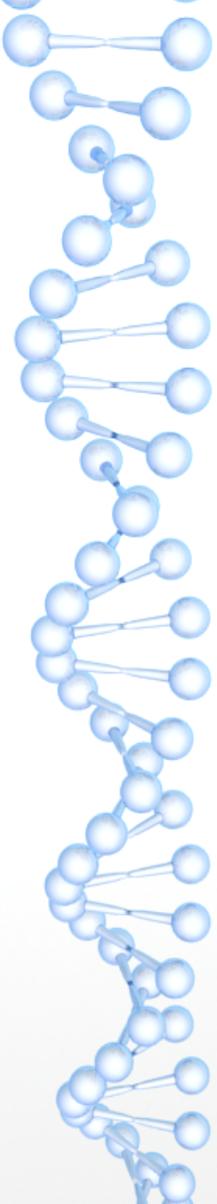
**DEAD OR ALIVE
REWARD
\$4,000**

- **Documentare** le pratiche di **modellazione** (trascrizione) e 'normalizzazione' del proprio progetto
 - Cosa sto trascrivendo? (Grafemi/allografi)? Come?
 - Cosa corrisponde a cosa (f/s)?
- Documentazione
 - In prosa ('specifiche' per futuri programmatori)
 - **Formale** (codice software, **tabelle**)



**La tabella dei segni
(T. Orlandi)**

La tabella dei segni



OCR/HTR
(testimone A)



Trascrizione manuale
(testimone B)

La tabella dei segni



Trascrizione
allografica

Vnder τhis Castle



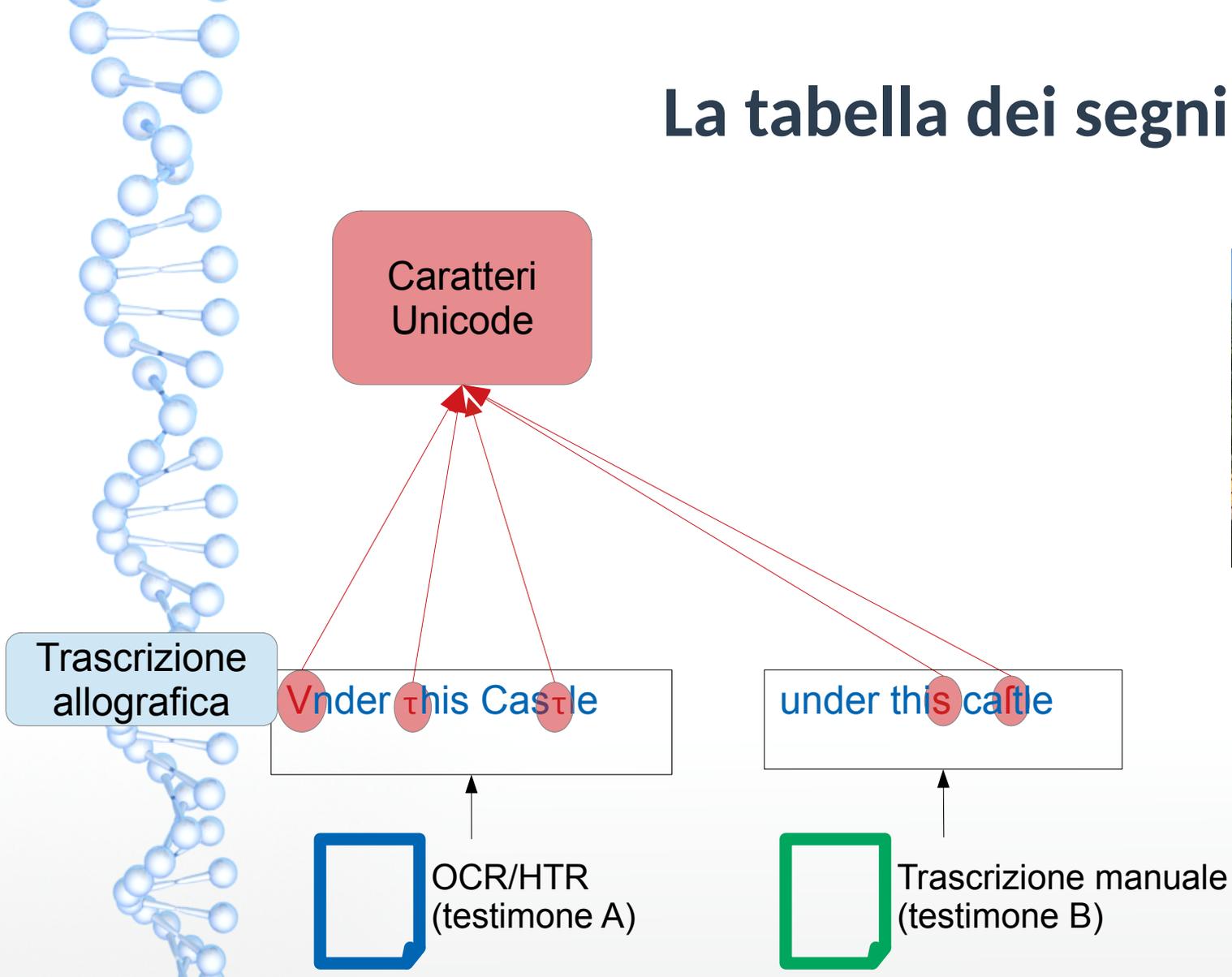
OCR/HTR
(testimone A)

under this caſtle



Trascrizione manuale
(testimone B)

La tabella dei segni



La tabella dei segni



Trascrizione
allografica

Vnder this Castle

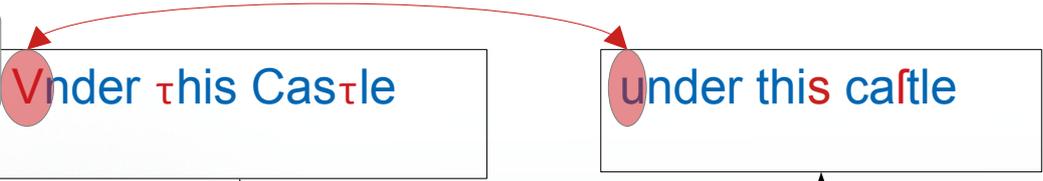
under this caftle



OCR/HTR
(testimone A)



Trascrizione manuale
(testimone B)



La tabella dei segni



Trascrizione
allografica

Vnder this Castle

under this caftle



OCR/HTR
(testimone A)



Trascrizione manuale
(testimone B)



La tabella dei segni



Trascrizione
allografica

Vnder this Castle

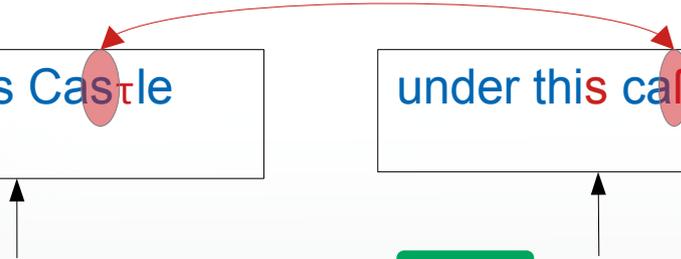
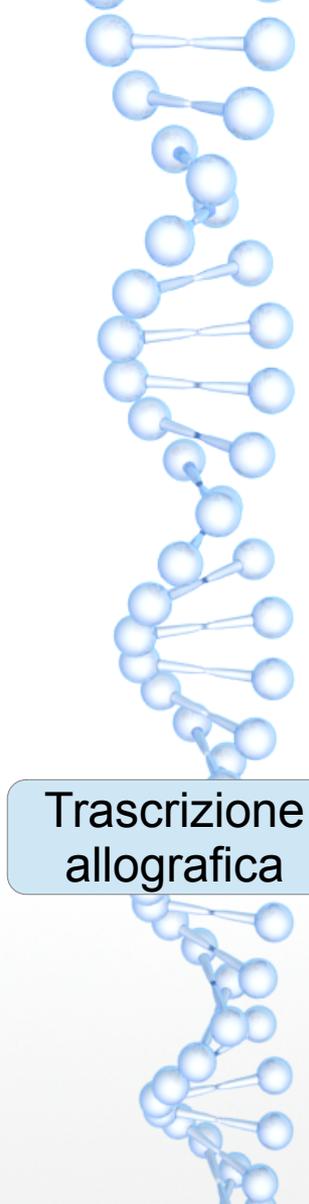
under this castle



OCR/HTR
(testimone A)



Trascrizione manuale
(testimone B)



La tabella dei segni



Trascrizione
allografica

Vnder τhis Castle



OCR/HTR
(testimone A)

under this caſtle



Trascrizione manuale
(testimone B)

La tabella dei segni



Gr	Allogr
s:	s
t:	τ ε √
u:	u ∨

Gr	Allogr
s:	s ʃ
t:	t
u:	u

Vnder τhis Castε

under this caʃtle

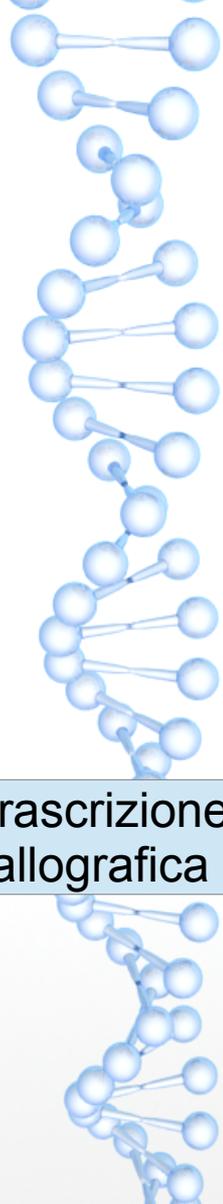


OCR/HTR
(testimone A)

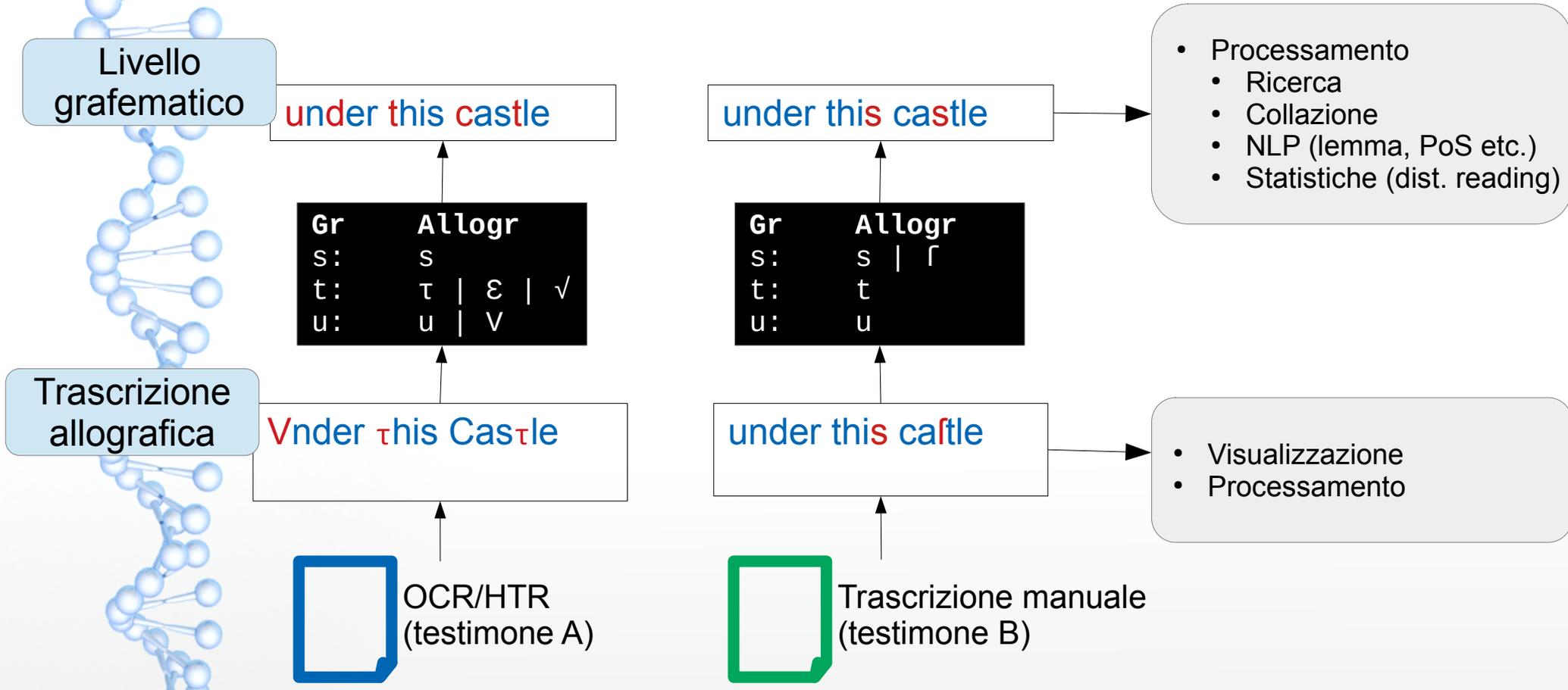


Trascrizione manuale
(testimone B)

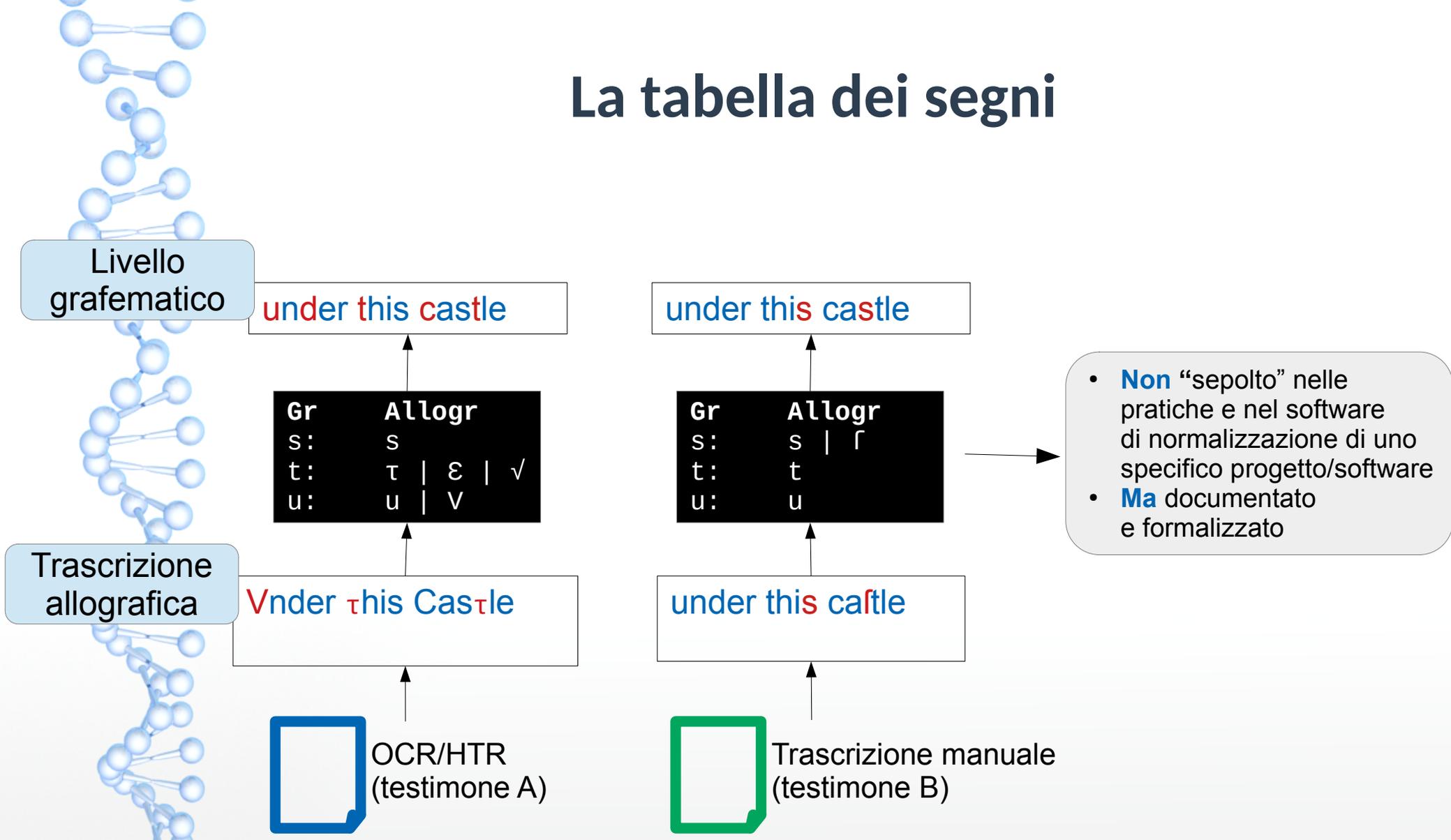
Trascrizione
allografica



La tabella dei segni



La tabella dei segni



**Processabilità,
modellazione,
documentazione**

Processabilità, modellazione, documentazione



WANTED

**DEAD OR ALIVE
REWARD
\$4,000**

- **Documentare** le pratiche di **modellazione** (trascrizione) e 'normalizzazione' del proprio progetto
 - Cosa sto trascrivendo? (Grafemi/allografi)? Come?
 - Cosa corrisponde a cosa (f/s)?
- Documentazione
 - In prosa ('specifiche' per futuri programmatori)
 - **Formale** (codice software, **tabelle**)

Processabilità, modellazione, documentazione

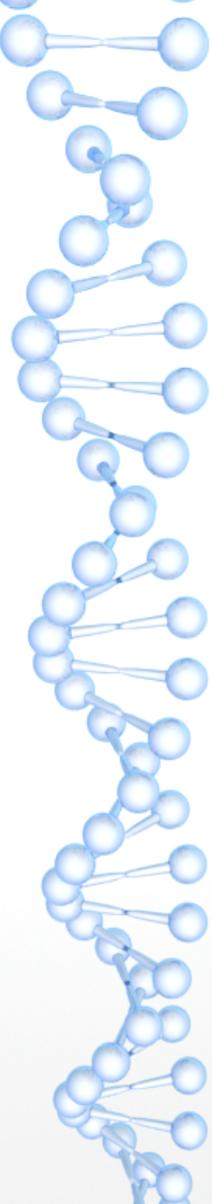


WANTED

**DEAD OR ALIVE
REWARD
\$4,000**

- Discorso scientifico sulla **modellizzazione**
 - Cos'è un grafema? Cosa un allografo?
- Modelli **condivisi**
 - Approccio 'dal basso', dal confronto
- Librerie **software** riusabili

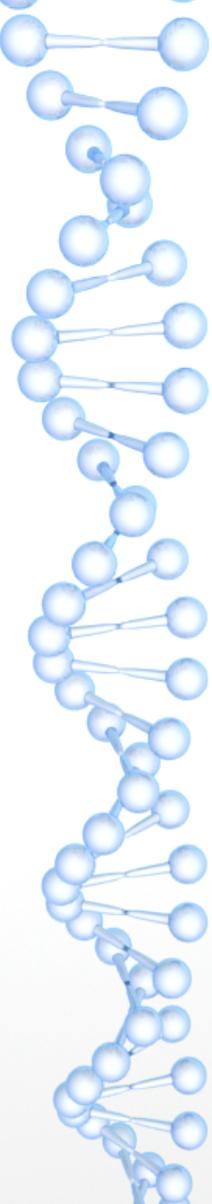
**Alcuni esempi
di tabella dei segni**



Alcuni esempi di tabella dei segni

- Vedi sezione [Seminar readings](#) (GitHub)
 - Human-readable tables of signs
 - Machine-readable tables of signs

Sommario



Sommario

- La **trascrizione diplomatica** delle fonti primarie
- **Processamento**
 - La questione
 - Le attuali soluzioni
 - Processabilità, modellazione, documentazione
- La **tabella dei segni** (T. Orlandi)
 - Alcuni esempi di tabella dei segni