

# Livelli di rappresentazione del testo nell'edizione del *De nomine* di Orso Beneventano

Paolo Monella, Università di Palermo

## 1 Introduzione

La comunicazione discute l'impianto metodologico e tecnologico dell'edizione critica digitale del *De nomine* di Orso Beneventano (<http://www.unipa.it/paolo.monella/ursus/transcription.html>), un trattatello grammaticale latino del IX secolo di ambiente beneventano (vd. Morelli 1910 e Fioretti 2010).

Si tratta di un prototipo che intende mettere alla prova ed approfondire le innovazioni metodologiche proposte da Tito Orlandi, sistematizzate in Orlandi 2010.

## 2 Metodo

La codifica XML-TEI è strutturata intorno ad elementi `<w>`, al cui interno il markup relativo alla trascrizione di fonti primarie (in particolare elementi come `<abbr>` e `<expan>`) subisce una specificazione semantica fondata sulla semiotica, in particolare la grafologia, e sui recenti sviluppi della paleografia digitale (Stokes 2011). L'obiettivo è codificare il testo del documento a più livelli testuali, semioticamente fondati. I livelli scelti per questa edizione sono:

1. *Livello grafematico*, le cui unità minime sono i grafemi specifici del sistema grafico del testimone, inclusi i segni paragrafematici e le brachilogie sistematiche. Tutti i caratteri Unicode contenuti in `<w>` rappresentano grafemi. Tutti i grafemi (glifi aventi valore distintivo) individuati dall'editore sono elencati e commentati nella "Graphemic Table of Signs" (GToS), un file CSV distinto dalla trascrizione XML-TEI ma che costituisce parte integrante dell'edizione. Tramite la GToS l'editore

fornisce una sua descrizione del sistema grafematico del manoscritto. La GToS rappresenta un'altra innovazione metodologica, realizzazione dell'idea di "tabella dei segni" teorizzata da Orlandi 2010 e giustificata dal noto principio saussuriano per cui un segno si definisce solo all'interno di un sistema semiotico, in contrasto con ogni altro segno di quel sistema. Se un manoscritto M non ha una distinzione tra "u" e "v" (dunque ha un solo glifo per entrambi, "a forma di u") ed un manoscritto N *ha* tale distinzione (dunque ha due glifi separati), il glifo "u" in M non è lo stesso grafema del glifo "u" in N, anche se "corrisponde" ad esso, né il glifo "v" in M è del tutto un altro grafema, in quanto anch'esso gli "corrisponde". Le pratiche attuali, ed anche le linee guida XML-TEI, si accontentano di usare lo stesso carattere Unicode "u" (U+0075) nella trascrizione grafematica del glifo "a forma di u" di M e di quello di N. Ciò è sufficiente per la visualizzazione a schermo e l'intuito del lettore, ma per ogni ulteriore elaborazione dei dati da parte del computer ciò implica fallacemente che "u" di M e "u" di N siano lo stesso grafema, mentre "u" di M e "v" di N sono grafemi diversi. La GToS serve dunque a "de-finire" ogni grafema all'interno del sistema grafico di uno specifico manoscritto. Appare però evidente che la collazione tra manoscritti diversi debba avvenire ad un livello di trascrizione più alto, cioè al livello alfabetico (se si vogliono confrontare le ortografie) o al livello linguistico (se interessano le sole varianti "significative").

2. *Livello alfabetico*, le cui unità minime sono lettere alfabetiche, qui chiamate "alfabemi" e distinte dai grafemi: i grafemi "significano" alfabemi. La GToS riporta, per ogni grafema, il suo significato alfabetico standard (ad es.: il significato standard del grafema codificato col carattere Unicode "u" è l'alfabema codificato col carattere Unicode "u", mentre il significato standard del grafema codificato con "ç" è la sequenza di alfabemi codificati rispettivamente con "q", "u" e "i"). La trascrizione al livello alfabetico non è codificata esplicitamente dall'editore per tutto il testo. Si danno infatti due casi: (A) Se un determinato grafema nel testo ha il suo significato alfabetico standard, indicato nella "Graphemic Table of Signs" (GToS), uno script JavaScript desume l'alfabema (o gli alfabemi) corrispondenti sulla base della trascrizione grafematica XML-TEI (file `casanatensis.xml`) e della GToS (file `GToS.csv`); (B) Se un determinato grafema non ha un significato alfabetico standard desumibile dalla GToS (ad es. l'abbreviazione iniziale in `c̄fer` per "confer", o il brevigrafo iniziale di "çā" per "quia"), l'editore, in fase di codifica, marca

il passaggio tramite un elemento `<choice>` con all'interno un `<abbr>`, che contiene una sequenza di grafemi, e un `<expan>` che contiene una sequenza di alfabemi, cioè il significato alfabetico di quei grafemi. Nel caso B, dunque, il software trova i due livelli (grafematico e alfabetico) già codificati esplicitamente. In sintesi, la codifica a livello alfabetico è fornita esplicitamente dall'editore all'interno di un elemento `<expan>` solo quando essa non sia desumibile dalla codifica grafematica tramite la tabella di corrispondenze fornita nella GToS.

3. *Livello linguistico*, le cui unità minime sono parole flesse, ma intese sinteticamente, non come sequenza di lettere alfabetiche. La parola flessa (es.: lat. “lupi”) viene così identificata nell'edizione tramite la combinazione di un lemma (“lupus, -i”) e di informazioni morfologiche (genitivo singolare). Nel codice XML-TEI, si sono usati a questo fine gli attributi dell'elemento `<w>` `@lemma` e `@ana` (per l'analisi morfologica). Nell'esempio di “lupi”, dunque: `<w ana="11B---B1--1" lemma="lupus" n="lupi">`. Nella realizzazione pratica dell'edizione, si è partiti da una trascrizione “normalizzata” di ogni parola, su cui si è fatto girare il lemmatizzatore/PoS tagger *TreeTagger* (<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>), col parameter file *Latin-ITTB UD treebank*, scaricato il 15/01/2016 dal sito *Universal Dependencies* (<http://universaldependencies.org/>) e basato sull'*IT-TB -Index Thomisticus Treebank* (<http://itreebank.marginalia.it/>). Il risultato della lemmatizzazione è stato poi verificato e corretto dall'editore. Attraverso la distinzione, in fase di codifica, del “livello linguistico” dai livelli “grafematico” ed “alfabetico”, la linguistica computazionale torna ad essere, come è sempre stata la linguistica in ambito tradizionale, supporto fondamentale dell'attività ecdotica.

### 3 Discussione

Quale di queste trascrizioni è dunque il “testo da leggere”? Quale corrisponde all'edizione critica tradizionale?

Uno dei vantaggi dell'edizione critica digitale è, in realtà, proprio quello di integrare in un sistema unico l'edizione cosiddetta “diplomatica” e quella “critica” o interpretativa, superando tale dicotomia (Vanhoutte 2000, Bodard and Garcés 2009, Orlandi 2010, Pierazzo 2011, Brüning, Henzel, and Pravida 2013, Pierazzo 2015, *Vercelli Book Digitale* <http://vbd.humnet.unipi.it/>, *Menota -Medieval Nordic Text Archive* <http://menota.org/>).

Nel modello di edizione qui proposto, se l'edizione “diplomatica” è facil-

mente identificabile con il livello grafematico, le funzioni svolte dall'edizione "critica" cartacea vengono distribuite a più livelli.

Nel caso di una tradizione testuale unitestimoniale come quella del *De nomine* di Orso Beneventano, la funzione-lettura viene diffratta tra il livello alfabetico e quello linguistico.

Il lettore leggerà il livello alfabetico se vuole ignorare le abbreviazioni e la punteggiatura antica (livello grafematico), ma è interessato all'ortografia originale del testo, senza alcuna "normalizzazione" delle monottongazioni o dell'ortografia antica, né correzione dei *lapsus calami*.

Leggerà invece il livello linguistico se è interessato al "testo" astratto: vi troverà l'ortografia e la punteggiatura "normalizzate", moderne, e i *lapsus calami* corretti. A questo livello, infatti, viene trascritta *la parola latina* che, secondo l'editore, lo scriba *intendeva scrivere*.

Se invece, a parere dell'editore, lo scriba intendeva effettivamente scrivere una determinata parola flessa (ad es. l'accusativo di un sostantivo), ma l'editore vuole *emendare* il testo trådito (ad es. in un dativo), questo tipo di intervento apparterrà anch'esso concettualmente al livello linguistico, ma andrà indicato tramite markup specifico. Per questa via, di fatto, si aggiunge un ulteriore livello di codifica del testo: il testo emendato dal filologo.

Un ulteriore livello di complessità sarà aggiunto, in futuri esperimenti di questo modello di edizione, nelle edizioni di testi con più testimoni.

In questi casi, ogni manoscritto avrà una sua edizione sui tre livelli citati, e la collazione in vista della definizione di un "testo dell'editore" confronterà le parole flesse al livello linguistico, evidenziando così solo le varianti dette "significative", ignorando invece le varianti "grafiche" (paleografiche, appartenenti al livello grafematico, o ortografiche, appartenenti al livello alfabetico).

Tale distinzione di livelli permette dunque di codificare univocamente la parola flessa al livello linguistico al di là del sistema grafico e persino dell'alfabeto utilizzati nei singoli testimoni, offrendo così un livello di codifica utile per la visualizzazione "normalizzata" (la tradizionale edizione "critica" o interpretativa), per la ricerca e l'analisi testuale, per l'interoperabilità dell'edizione al di là delle specificità paleografiche del manoscritto e, potenzialmente, per la generazione di apparati critici dinamici in tradizioni pluritestimoniali.

## 4 Sviluppo, licenze e riuso

L'edizione è *open source*: tutto il codice sorgente (markup e software) è stato scritto con l'editor Vim (<http://www.vim.org/>) ed è disponibile sul

repository <https://github.com/paolomonella/ursus> sotto la GNU General Public License, insieme ad una ampia documentazione in inglese.

La scelta dei linguaggi utilizzati, e in particolare di XML-TEI, è mirata a consentire il riuso del codice sorgente. Allo stesso fine mira la documentazione dettagliata di ogni aspetto, metodologico e tecnologico, del sistema-edizione, e in particolare della semantica degli elementi XML-TEI utilizzati.

L'impianto teorico dell'edizione prevede che le riproduzioni digitali del *codex unicus* che contiene il testo (Casanatensis 1086) siano collegate alla trascrizione. Tuttavia, un accordo con la Biblioteca Casanatense di Roma, che ha fornito le riproduzioni, attualmente vincola l'editore a non rendere disponibili le immagini online.

## References

- Bodard, Gabriel, and Juan Garcés. 2009. "Open Source Critical Editions: A Rationale". In *Text Editing, Print, and the Digital World*, ed. by Marilyn Deegan and Kathryn Sutherland, 83–98. Digital Research in the Arts and Humanities. Aldershot: Ashgate.
- Brüning, Gerrit, Katrin Henzel, and Dietmar Pravida. 2013. "Multiple encoding in genetic editions: the case of "Faust"". *Journal of the Text Encoding Initiative*, no. 4. doi:10.4000/jtei.697.
- Fioretti, Paolo. 2010. "L'eredità di un maestro. Genesi ed edizione della grammatica di Orso beneventano". In *Libri di scuola e pratiche didattiche. Dall'antichità al Rinascimento, Atti del Convegno Internazionale organizzato dall'Università degli Studi di Cassino (Cassino 7-10 maggio 2008)*, ed. by Lucio Del Corso and Oronzo Pecere, 293–328. Cassino: Edizioni Università di Cassino.
- Morelli, Camillo. 1910. "I trattati di grammatica e retorica del cod. Casanatense 1086. Nota del dott. C. Morelli, presentata dal Socio G. Vitelli". *Rendiconti della Reale Accademia dei Lincei, Classe di scienze morali, storiche e filologiche* XIX:287–328.
- Orlandi, Tito. 2010. *Informatica testuale. Teoria e prassi*. Roma: Laterza.
- Pierazzo, Elena. 2011. "A rationale of digital documentary editions". *Literary and Linguistic Computing* 26, no. 4 (): 463–477. doi:10.1093/llc/fqr033.
- . 2015. *Digital Scholarly Editing: Theories, Models and Methods*. Farnham, Surrey, UK and Burlington, VT: Ashgate.

- Stokes, Peter A. 2011. "Describing Handwriting, Part IV: Recapitulation and Formal Model". *DigiPal Blog* October 14 (). <http://www.digipal.eu/blogs/blog/describing-handwriting-part-iv/>.
- Vanhoutte, Edward. 2000. "Where is the editor? Resistance in the Creation of an electronic critical edition". In *DRH 98. Selected papers from Digital resources for the Humanities 1998*, ed. by Marilyn Deegan, Jean Anderson, and Harold Short, 171–183. London: Office for Humanities Communication. <http://etjanst.hb.se/bhs/ith//1-99/ev.htm>.