

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

Abstract esteso

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

Il testo è un trattato grammaticale in latino del IX secolo di ambiente beneventano. 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 separato 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 (vd. Orlandi 2010).
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 "ϕ" 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 "ċfer" per "confer", o il brevigrafo iniziale di "ċa" 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* col parameter file *Latin-ITTB UD treebank*, scaricato il 15/01/2016 dal sito *Universal Dependencies* e basato sull'*Index Thomisticus Treebank* (IT-TB). Il risultato della lemmatizzazione è stato poi verificato e corretto dall'editore. Per questo verso, la linguistica computazionale torna ad essere, come è sempre stata la linguistica in ambito tradizionale, supporto fondamentale dell'attività ecdotica. Ciò permette di codificare univocamente la parola flessa 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 e per l'interoperabilità dell'edizione al di là delle specificità paleografiche del manoscritto. Un'ulteriore applicazione di questa distinzione di livelli è, in edizioni di tradizioni pluritestimoniali (cioè non in questo caso) la possibilità di effettuare una collazione tra manoscritti che evidenzia solo le varianti dette "significative", ignorando invece le varianti "grafiche" (paleografiche e ortografiche). Il modello qui

sperimentato intende proporre una soluzione alla questione aperta del rapporto fra edizione diplomatica e critica (vd. *Vercelli Book Digitale, Menota*, Brüning et al. 2013), superando la vecchia dicotomia grazie ad un ripensamento profondo, semioticamente fondato, dei livelli di trascrizione e di edizione di un testo.

Bibliografia

Brüning, G., Henzel, K., Pravida, D. (2013), *Multiple Encoding in Genetic Editions: The Case of "Faust"*, *Journal of the Text Encoding Initiative* 4.

Fioretti, P. (2010), *L'eredità di un maestro. Genesi ed edizione della grammatica di Orso beneventano*, in Lucio Del Corso & Oronzo Pecere, ed., *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)*, Edizioni Università di Cassino, pp. 293-328.

Morelli, C. (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, T. (2010), *Informatica testuale. Teoria e prassi*, Laterza.

Pierazzo, E. (2016), *Digital Scholarly Editing: Theories, Models and Methods*, Taylor & Francis.

Stokes, P. A. (2011), *Describing Handwriting, Part IV: Recapitulation and Formal Model*, DigiPal blog, <http://www.digipal.eu/blog/describing-handwriting-part-iv-recapitulation-and-formal-model/>. Questo sito e i successivi citati in bibliografia sono stati consultati il 31/05/2016.

Index Thomisticus Treebank (IT-TB), <http://itreebank.marginalia.it/>.

Medieval Nordic Text Archive (Menota), <http://menota.org/>.

TreeTagger, <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>.

Universal Dependencies, <http://universaldependencies.org/>.

Vercelli Book Digitale, <http://vbd.humnet.unipi.it/>.