

DANIELE FUSI

Edizione epigrafica digitale di testi greci e latini: dal testo marcato alla banca dati

Obiettivi di questa edizione

Negli ultimi anni si vanno moltiplicando i progetti di edizioni elettroniche di materiale classico, sia esso letterario, papiraceo o epigrafico; in quest'ultimo campo, benché a partire dai fondamentali inquadramenti di Krummrey-Pancierera¹ siano stati definiti gli indispensabili criteri editoriali, le realizzazioni concrete di edizioni elettroniche e i semplici progetti appaiono ancora alquanto vari e diseguali, sia nel materiale preso in esame che nella sua strutturazione sul piano della tecnologia dell'informazione. Nondimeno, è probabile che ognuno di questi progetti indipendentemente dal suo livello di attuazione possa apportare un contributo alla definizione di una qualche base comune che renda possibile una loro futura integrazione, sicché nessuno degli sforzi compiuti per ogni progetto debba andare perduto o rimanere isolato nel suo ambito senza possibilità di dialogo con altri sistemi. Non si tratta qui di definire un modo univoco di strutturare e gestire l'informazione relativa ai testi epigrafici, compito difficilmente realizzabile e probabilmente discutibile, data l'infinita varietà di intenti, strumenti e prospettive di ricerca che ogni progetto può voler realizzare nel suo ambito specifico; piuttosto, interessa offrire un contributo per illuminare i diversi punti di vista relativi a un oggetto che proprio nella varietà dei suoi aspetti riguarda da vicino diversi campi di studio al di là dell'epigrafia in senso stretto: archeologia, paleografia, storia, filologia, linguistica, etc. In questo senso ogni progetto può suggerire aspetti interessanti a una visione più comprensiva dell'insieme, e contribuire in prospettiva alla definizione di una base comune capace di consentire almeno lo scambio dei dati fra sistemi diversi, sì da rendere possibile una integrazione dei vari *corpora* elettronici la cui utilità è anche direttamente proporzionale alla loro completezza.

In questo ambito, il progetto qui presentato nasce anzitutto nella prospettiva del filologo e dell'informatico, ed è fortemente improntato da una decisa separazione tra forme e contenuti, basata sulla nozione chiave della trasformazione: un unico contenuto, altamente strutturato dal punto di vista semantico, può assumere forme potenzialmente infinite, sia riguardo alla selezione e all'ordinamento del materiale che al suo formato elettronico. La trasformazione di questo contenuto si adatterà di volta in volta ai diversi tipi di pubblico dell'opera (che può variare dal visitatore occasionale di un museo allo studente fino allo studioso di professione) come ai diversi *media* in cui essa può essere pubblicata (*cd-rom* o *dvd*, siti *web*, libri cartacei o elettronici, chioschi multimediali, etc.). Inoltre, il medesimo contenuto deve poter essere espandibile non solo in senso quantitativo (con l'aggiunta di nuovo materiale), ma anche in senso qualitativo, consentendo cioè l'aggiunta all'informazione preesistente di informazione specializzata, magari nel senso di una specifica disciplina di studi. Quest'ultimo requisito risulta particolarmente rilevante specie in ambito epigrafico, dove la quantità di materiale e il suo interesse per numerose altre discipline (filologia, linguistica, storia, archeologia, paleografia, etc.) è tale da rendere indispensabile ottimizzare le risorse investite in un'edizione, consentendo in massimo grado il riuso dell'informazione esistente e la sua adattabilità in qualsiasi particolare direzione. L'edizione progettata deve quindi essere anzitutto unica per tutti i *media*, pubblici e diversi campi di specializzazione cui ci si desidera riferire, consentendo così una sua gestione centralizzata indipendente dalla sua pubblicazione in un ambito specifico: in tal modo si minimizza lo sforzo di manutenzione dell'informazione, che rimane unica benché votata a diverse edizioni. Ciò naturalmente implica che l'informazione risulti quanto più possibile autonoma e astratta da una qualsiasi veste specifica, che finirebbe inevitabilmente per condizionare i suoi impieghi e limitare la sua espandibilità: in questo senso, la netta separazione fra l'unico contenuto e le diverse forme in cui potenzialmente esso si può incarnare richiede un sistema profondamente votato alla trasformazione dinamica della sua informazione, ovvero un'attenta progettazione della sua struttura e la creazione di numerosi strumenti *software* realizzati *ad hoc* per gestire immissione, modifica e conversione dei dati in qualsiasi direzione. Non si tratta quindi di assumere a modello alcuna specifica forma finale di edizione e sulla sua base arricchire un testo epigrafico della necessaria informazione metatestuale, ad esempio imitando una scheda epigrafica con un testo variamente marcato da *tag*: piuttosto, si intende creare qui una vera banca dati nel senso più generico del termine, dove l'informazione è strutturata in modo da isolare ogni elemento semanticamente diverso in un suo contenitore separato, con una sorta di scomposizione della tradizionale scheda epigrafica fino agli atomi dei suoi componenti costitutivi. Questa informazione elementare così atomizzata e classificata potrà poi essere ricomposta in modo automatico e dinamico in

¹ H. Krummrey - Silvio Panciera, *Criteri di edizione e segni diacritici*, «Tituli» 2 (1980) 205-215.

qualsiasi forma specifica richiesta dal particolare tipo di edizione, formato elettronico, *medium* e pubblico cui ci si intenda riferire.

La trasformazione come dote intrinseca del progetto editoriale è inoltre parte dell'indispensabile attenzione alla longevità dell'opera che si intende realizzare: non è infatti raro il caso di opere digitali concepite solo pochi anni fa eppure già rese obsolete non solo dall'avanzare delle tecnologie, fatto comunque inevitabile sul medio e lungo periodo, ma anche e soprattutto dalla loro eccessiva peculiarità nella strutturazione e codifica del testo. Persino nel campo dell'editoria commerciale infatti può accadere che la versione digitale di opere stampate pochi anni fa risulti oggi addirittura indecifrabile costringendo a complessi sforzi di analisi e riconversione, perché realizzata magari con l'ausilio di codifiche proprietarie ideate *ad hoc* per quella data opera e gli strumenti informatici dell'epoca. Se non è dunque possibile (né d'altronde auspicabile) che le tecnologie impiegate oggi rimangano immutate nonostante i rapidissimi mutamenti dell'orizzonte informatico, è però necessario adoperare ogni sforzo nell'ideazione e realizzazione di opere digitali di portata così ampia come quelle di testi epigrafici o letterari, in modo che in un futuro non troppo remoto sia possibile riconvertirle facilmente piuttosto che abbandonarle all'obsolescenza. A questo scopo, accanto alla capacità di trasformazione e alla complementare tendenza alla massima separazione possibile dell'unico contenuto dalle sue molteplici forme occorre basare l'opera digitale su standard ampiamente diffusi e affermati, per i quali sussiste la garanzia di un'adeguata documentazione e dotazione di *software* capace di trattarli per gli scopi dell'edizione, come in futuro di convertirli in altre tecnologie.

Specificità dell'edizione digitale

In termini più semplici si può meglio intendere la natura di questo progetto a partire da alcuni banali esempi concreti: nel campo dell'editoria elettronica specie italiana spesso accade che opere tradizionalmente edite su carta ricevano una veste elettronica di solito ancillare rispetto alla loro veste originaria, come ad esempio per dizionari, repertori testuali o altre opere di consultazione. Tuttavia non è raro constatare come in questi casi l'edizione elettronica finisca per ridursi a un mero clone di quella cartacea, trasposta più o meno identica dal *medium* del libro tradizionale a quello del *cd-rom* o *dvd*: accade persino che si offra la semplice versione elettronica del testo stampato, aggiungendo magari qualche funzione di ricerca a tutto testo o qualche contenuto multimediale, senza tenere in alcun conto la specificità del *medium* elettronico.

Un testo stampato infatti necessariamente offre solo un limitato numero di modi di accedere all'informazione in esso contenuta, che viene predisposta dall'autore in un certo ordine, con una certa selezione e una specifica veste tipografica, magari aggiungendo a corredo diversi indici per la consultazione sì da consentire un uso più diversificato dell'opera. Per quanta cura si possa mettere in tali operazioni è comunque inevitabile che il ventaglio degli usi consentiti da una simile opera rimanga intrinsecamente limitato dalla forma in cui essa si è incarnata per la stampa. Per fare un esempio banale, l'edizione di un *corpus* testuale può contenere un indice di antroponimi, toponimi, o altri aspetti ritenuti notevoli per una qualsiasi disciplina (letteratura, storia, geografia, prosopografia, archeologia, religione, linguistica, etc.), consentendo così di fruire dell'opera in modi diversi corrispondenti agli intenti del suo utente; ma quando un particolare tipo di indice manchi lo strumento offerto da tale edizione sarà di fatto praticamente inutile, a meno di non effettuare una faticosa ricerca manuale al suo interno, tanto meno fattibile quanto più l'opera risulti voluminosa. In tal caso l'utente potrà rivolgersi magari a edizioni specializzate, laddove esse siano disponibili, o integrare l'informazione incrociando i dati provenienti da altre edizioni, ma è comunque evidente che tali operazioni richiederanno molto tempo, quando pure esse risultino possibili in base al materiale disponibile.

In questo senso, un'edizione elettronica che si limiti a riprodurre *in toto* quella cartacea, o in altri termini la stessa edizione cartacea trasportata su un *medium* diverso, può certo offrire un minimo vantaggio consentendo se non altro ricerche a tutto testo sull'intero contenuto dell'opera, allo stesso modo in cui è ovviamente più facile trovare tutte le occorrenze di una parola in un testo digitale piuttosto che sfogliando le pagine della sua forma stampata. Tuttavia è facile comprendere come una simile trasposizione dei contenuti in un altro *medium* non costituisca propriamente una vera opera digitale, nella misura in cui essa condivide con la cartacea le stesse limitazioni strutturali: che sia infatti più semplice cercare una parola al suo interno dipende semplicemente dalla diversità del *medium*, ma rimane il fatto che entrambe le edizioni mostrano unici ordinamenti, selezioni, indicizzazioni e forme di presentazione dei dati, che l'utente non può in alcun modo alterare a posteriori: in altri termini, entrambe le edizioni in sé dicono sempre la stessa cosa e lo fanno sempre nello stesso modo, secondo la vecchia accusa mossa da Platone alla parola scritta. Se poi si pensa a testi più complessi come vocabolari o anche schede epigrafiche è ovvio che anche la ricerca a tutto testo perderà molta della sua utilità, applicandosi indistintamente a una lunga e amorfa sequenza di parole, senza possibilità di distinguere cosa al suo interno rappresenti nel caso di un vocabolario lemma, traducevole, citazione,

paradigma, spiegazione, etimo, etc., o nel caso di una scheda epigrafica classificazione, collocazione storica e geografica, testo, apparato, commento, etc. La consultazione di un tale vocabolario elettronico diverrebbe quindi pressoché infruttuosa, dato che non sarebbe possibile limitare la ricerca a un ambito semantico specifico (ricerca per lemma, ricerca per traduce, ricerca per citazione, etc.) all'interno dell'indistinto flusso di parole che costituisce il suo testo. Qualora invece in questo flusso indistinto di parole si individuino in modo univoco ogni porzione rilevante di testo secondo il proprio ruolo semantico (quanto rappresenta il lemma, l'etimologia, la morfologia, il traduce, la fraseologia, etc.) diverrà subito possibile utilizzare questo unico contenuto per molteplici scopi e in molteplici forme: ad esempio si potranno generare automaticamente elenchi alfabetici (diretti o inversi) dei lemmi, o raggruppamenti di lemmi per categoria morfologica o attestazione in una data opera, autore o contesto, o anche realizzare sempre in modo del tutto automatico un rudimentale vocabolario bilingue, semplicemente estraendo tutti i traduce in un elenco alfabetico con il rimando al relativo lemma. L'utente dell'edizione digitale potrà inoltre effettuare ricerche combinate anche molto complesse, ad esempio per trovare tutti i lemmi iniziati con determinate lettere, appartenenti a determinate categorie morfologiche, riconducibili alla medesima radice, attestati in determinati autori e/o opere, etc.: è insomma evidente che in tal modo le possibilità di consultazione di una simile edizione risulteranno subito virtualmente infinite, fornendo uno strumento assai più utile di una mera trasposizione digitale di un'opera concepita su carta. Può certo accadere che siano di fatto edite come opere elettroniche mere trasposizioni su *cd* o *dvd* di testi pubblicati a stampa, ignorando completamente le specificità del diverso *medium*, per mancata strutturazione dei dati preesistenti, scarsa capacità o convenienza di affrontare una loro vera riconversione elettronica, o fraintendimento della specifica natura dell'edizione elettronica rispetto alla cartacea; tuttavia la specificità dell'edizione digitale consiste proprio nelle possibilità virtualmente infinite di consultazione e presentazione dei dati che essa offre. Una volta infatti che si disponga di un'informazione di partenza opportunamente strutturata sul piano semantico, diviene possibile selezionarla, riordinarla e assegnarle la veste tipografica più adatta alle puntuali richieste che l'utente pone di volta in volta al programma di consultazione: in questo senso non si ha più una pubblicazione, ma di fatto il materiale viene ripubblicato dinamicamente in tempo reale a ogni specifica richiesta dell'utente, nei modi e nelle vesti più adatti a soddisfare tale richiesta, incrociando magari i dati secondo modalità complesse dipendenti dai parametri forniti dall'utente. Per fare un esempio banale, chiunque ha sperimentato quanto differiscano una ricerca di una parola in un indice di un libro e la ricerca di termini su un motore di ricerca in *internet*, con la possibilità di incrociare i parametri, associarli con operatori, restringere o allargare i campi di interrogazione, effettuare nuove ricerche a partire da risultati appena ottenuti, etc.².

Peculiarità del documento epigrafico

Il documento epigrafico da cui trae origine il progetto qui presentato è ovviamente peculiare nella misura in cui è portatore di un testo giuntoci direttamente sul suo supporto originale. Qualsiasi edizione deve quindi tener conto di questo duplice aspetto materiale e linguistico: il testo è parte integrante di un monumento inserito in un dato contesto, sicché l'aspetto materiale (manufatto, contesto, decorazione, funzione, etc.) deve sempre potersi affiancare ad esso, che è prezioso testimone dell'evoluzione della lingua o delle sue particolarità regionali³. Si consideri ad esempio l'epigrafe greca riportata nella fotografia (digitale) qui sotto accanto alla sua edizione cartacea (in IGUR 443):

² In realtà un tale esempio per quanto universalmente comprensibile potrebbe risultare fuorviante nella misura in cui di fatto le attuali tecnologie di ricerca di informazione su *web* si basano su indicizzazioni più o meno complesse ma comunque puramente formali dei contenuti delle pagine HTML, come chiunque sperimenta continuamente nel confrontarsi con la ridondanza e la pertinenza dei risultati ottenuti rispetto alle proprie intenzioni. Probabilmente solo OWL o altre tecnologie di definizione semantica dei contenuti potranno fornire la base per ricerche veramente mirate e infinitamente più efficaci delle attuali.

³ Ovviamente il lapicida che nota ad esempio con I quanto oggi si scrive con EI non fa altro che esprimere in forma scritta un suono della propria lingua corrente. Come avvertiva Reinach, pure con qualche schematismo (1885 p.238), "le mot d'*ortographe* éveille en nous une idée de règle qui a été longtemps étrangère à l'antiquité. Pour nous, l'*ortographe* est une manière fixe d'écrire les mots, en dépit souvent de la prononciation qu'on leur donne; pour les anciens jusqu'à l'époque alexandrine [...], l'*ortographe* proprement dite n'existe pas et l'on écrit les mots comme on les prononce. L'écriture était vivante chez eux; elle est savante chez nous. Ils la conformaient à la parole; nous la réglons sur l'étymologie ou la tradition. Chez eux, elle représente ce qui se dit; chez nous, elle est souvent un témoignage de ce qui ne se dit plus".



443 Cippus ex marmore altus 0,89, latus 0,41, crassus 0,215. Litterae 0,03 in initio; 0,015 in ll. 8-9. In S. Paulo extra muros descripserunt Muratori, *Thesaurus...*, p. 1443, 7 et Gori, *Marucell.* A 77 f. 90 et *Symb. Litter.* V, 24 missum a P. P. Ginannio; Donati, *schol. bibl. Lucens.* 1202 f. 50; Marini, *Vatic.* 9133 f. 253; male Mercklin, *Arch. Zeit.*, 7 (1849) p. 106 nr. 16. Ex editis Franz, *CIG* 6354. Denuo descripsit Kaibel, *JG* XIV 1503 et add. (inde Cagnat, *JGR* I 234). Cfr. Friedlaender, *Sittengeschichte*⁴, II, p. 35 n. 7; A. Neppi Modona, *Hommages à L. Herrmann*, 1960, p. 565-570. Ibidem contuli.

	Θ(εός)	Κ(αταχθονίος)	
		Βασιλείου	
		ΝΕΚΟΝ	χά-
		ριν	ἐπύη-
urceus	σεν	Μ(άρκος) Αὔρ(ήλιος)	patera
5		Διονύσις	
		ἀδελφῶν	
	ἰς	ἔθνος Βενετιανούς,	πρίν-
			κιπτι

Come si vede, dinanzi a grafie non standard come ἐπύησεν, Διονύσις, ἰς l'editore ha scelto di conformarsi al testo originale piuttosto che la normalizzazione ortografica ἐποίησεν, Διονύσιος, εἰς: naturalmente nel caso di un'edizione cartacea una scelta (a meno di non voler stampare il testo due volte) è d'obbligo, e qualsiasi scelta non può che risultare compromissoria: nel caso della grafia 'storica' si rispetta la storicità del documento originale ma si sacrifica la leggibilità dell'edizione per il pubblico meno specializzato; nel caso della grafia normalizzata si stampa un testo facilmente comprensibile da qualsiasi utente dell'edizione e conforme alle nostre convenzioni ortografiche, ma inevitabilmente falsato rispetto alla realtà linguistica coeva. È naturalmente possibile aggiungere un apparato con le lezioni originali, ma in ogni caso il testo che si presenta in prima istanza deve essere il frutto di una delle due scelte, e motivi pratici spesso limitano la possibilità di aggiunte di questo tipo a un'edizione su carta. Peraltro, questo stesso esempio mostra come una simile scelta possa risultare ancora più complessa: non si tratta solo di 'meccaniche' e relativamente banali trasposizioni ortografiche corrispondenti all'itacismo (οι per υ, ει per ι), ma di fenomeni di evoluzione della lingua dotati spesso di conseguenze importanti per l'intero sistema. Una forma come Διονύσις infatti rappresenta l'esito di una catena di trasformazioni linguistiche che, benché basata sullo stesso fenomeno di partenza (l'itacismo), finisce per avere conseguenze sul sistema morfologico configurandosi come un vero metaplasmo. È infatti noto che dall'età ellenistica la tendenziale caduta della vocale posttonica in una forma come Διονύσι(ο)ς provoca una generale deriva dei sostantivi in -(ο)ς appartenenti alla seconda declinazione, che terminando ormai in -is/ come i maschili della prima declinazione in -ης = -is/ finiscono per passare a quest'ultima, essendo una grafia come Ἀντώνης assai più comprensibile al parlante rispetto all'uscita -ις di una forma sincopata come Ἀντώνις⁴. In tal caso ci si trova dunque dinanzi a un vero metaplasmo che ha le sue origini nella fonetica e si ripercuote pesantemente sul sistema flessivo nominale della lingua greca, dinanzi al quale risulterà evidente che nessuna scelta editoriale si può imporre in modo sistematico. D'altronde rimane necessaria l'adozione di una convenzione ortografica omogenea almeno all'interno della medesima edizione, e il problema risulta ancor più importante se si pensa poi alla costruzione di indici e repertori, specie nel caso di edizioni digitali: è infatti ovvio che memorizzando solamente una forma come ἐπύησεν in una ipotetica edizione digitale l'utente che cercasse tutte le occorrenze della parola ἐποίησεν non troverebbe alcun risultato benché un testo presentante questa parola (sia pure in una forma diversa) sia presente nel *corpus*, e viceversa. Proprio un'edizione digitale però può offrire l'opportunità di superare le inevitabili prassi compromissorie legate per ovvi motivi pratici all'edizione cartacea, nella misura in cui rispettando la sua specificità diviene possibile presentare dinamicamente lo stesso testo in modi diversi, con tutti i dati complementari che si ritengono utili.

Realizzazione

La realizzazione di questo progetto prende dunque le mosse dal rispetto delle specificità dell'edizione digitale da un lato, e delle peculiarità del testo epigrafico dall'altro, con la sua duplice faccia materiale e testuale.

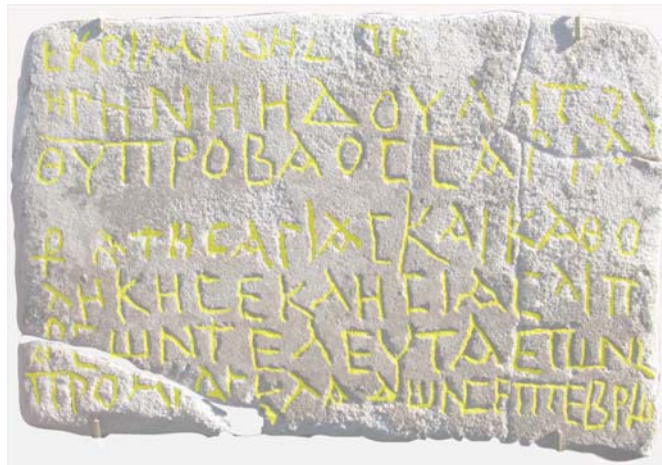
⁴ Alla fine del periodo romano lo stesso fenomeno assieme al proliferare dei diminutivi in -tov e forse all'influsso del latino *-arium* provoca la creazione di neutri in -iv (spesso -αiv) di seconda declinazione.

a) **Aspetto materiale: multimedialità, interattività, applicazioni grafiche**

La prima ovvia fase per un'edizione come quella qui presentata è costituita dall'acquisizione del materiale, tramite la realizzazione di fotografie digitali da un lato e il reperimento di tutta la principale bibliografia relativa dall'altro. Sin dall'inizio quindi l'edizione assume una fisionomia multimediale, ovviamente tanto più adatta a un'edizione digitale, e tale aspetto è riflesso anche nei percorsi paralleli che testo e immagini seguono nelle fasi successive. Da un lato dunque si procederà al trattamento elettronico del materiale fotografico, sottoposto ove necessario a più o meno complesse manipolazioni digitali su specializzati applicativi di fotoritocco (anzitutto *Adobe Photoshop*) allo scopo di isolare ed esaltare al meglio l'informazione testuale proveniente dalla pietra. Inoltre, un particolare aspetto di tali manipolazioni è rappresentato per ogni singola iscrizione (o solamente per quelle ritenute più rilevanti) dalla realizzazione di un suo fedele calco digitale, ottenuto in modo del tutto virtuale a partire dalla sua immagine: con la sovrapposizione di 'fogli' virtuali all'immagine fotografica originale e il tracciamento dei contorni delle lettere si ottiene un vero e proprio calco, tendenzialmente assai fedele ai tratti dell'originale, isolato poi dallo sfondo e ricolorato per essere proiettato su una superficie bianca, come un tradizionale disegno a china. Rimandando per qualche ulteriore dettaglio su queste fasi a un mio precedente intervento che sarebbe inutile ripetere in questa sede⁵, è sufficiente mostrare un piccolo esempio relativo allo stadio iniziale e finale dell'estrazione del calco digitale dalla fotografia opportunamente ritoccata:

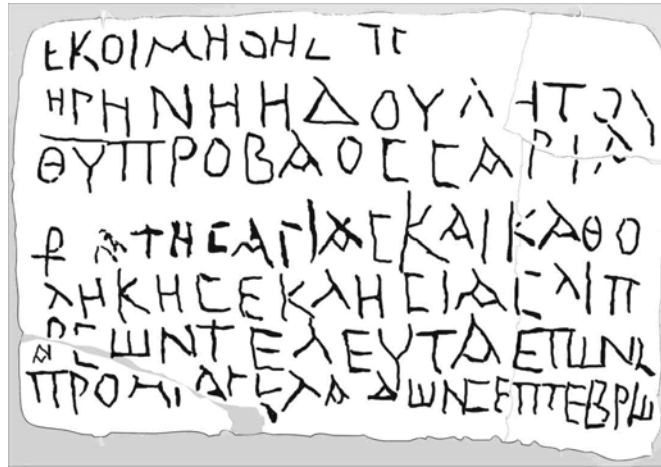


1) fotografia originale ritoccata



2) una delle fasi intermedie per la tracciatura del calco digitale

⁵ *Epigrafia informatica: progetto per un'edizione elettronica di materiale epigrafico*, «Seia» n.s. 6-7 (2001-2002) 177-210.



3) estrazione del calco digitale

Al termine di questo processo, per ogni epigrafe si disporrà così di una o più fotografie e uno o più calchi digitali da esse estratti. Parallelamente (e/o successivamente, in funzione del numero dei redattori assegnati al progetto) il testo dell'iscrizione verrà stabilito anche in rapporto ai risultati dell'esame delle fotografie e naturalmente a tutte le risorse bibliografiche utili, e immesso nella banca dati tramite un apposito programma.

Inoltre, il redattore può aggiungere contestualmente anche un insieme di informazioni sussidiarie alle immagini dell'iscrizione (fotografie e calchi), rappresentanti la connessione diretta fra il testo e la sua immagine grafica: per ogni immagine infatti è possibile definire una serie di 'mappe', strati, didascalie e misurazioni virtuali, implementati tutti in modo del tutto separato dall'immagine stessa in file XML distinti:

- le mappe correlano porzioni di testo dell'epigrafe a regioni della sua immagine, in modo che l'utente possa visualizzare entrambe in modo sincronizzato: il programma di consultazione consentirà infatti di raggiungere la porzione di testo desiderata nell'immagine, o viceversa di individuare la porzione di testo semplicemente posizionando il cursore sulla regione ad esso corrispondente nell'immagine stessa. Il redattore definisce visualmente le regioni di testo sull'immagine dell'epigrafe, e un programma provvede a generare il codice XML ad esse corrispondente, che sarà a sua volta interpretato dai programmi di consultazione per applicare una o più mappe a una data immagine. Ogni immagine può avere un qualsiasi numero di mappe, ciascuna deputata a evidenziare particolari regioni visuali al suo interno.
- gli strati sono immagini di varia natura sovrapponibili come strati successivi all'immagine fotografica di base dell'epigrafe: ad esempio, il calco viene tipicamente visualizzato come uno strato dell'immagine da cui è tratto, variando il suo grado di opacità in modo da lasciar trasparire in modo più o meno evidente l'immagine sottostante, sì da consentire all'utente l'immediato riscontro dei tracciati delle lettere. Ogni immagine può avere un qualsiasi numero di gruppi di strati, ciascuno deputato a evidenziare particolari aspetti grafici.
- le didascalie sono vere e proprie didascalie dal contenuto arbitrario, che possono essere visualizzate in un qualsiasi punto dell'immagine fotografica, ad esempio per evidenziare un qualsiasi aspetto visuale dell'epigrafe. Ogni immagine può avere un qualsiasi numero di gruppi di didascalie, ciascuno deputato a marcare un insieme coerente di osservazioni visuali.
- le misurazioni virtuali sono misurazioni di un qualsiasi elemento della pietra (tipicamente usate a scopi paleografici), effettuate virtualmente sulla fotografia digitale con riferimento alla scala metrica fotografata con l'iscrizione (vedi sotto).

Ancora una volta, in riferimento alla filosofia generale del progetto, è importante osservare come tutti questi elementi relativi all'immagine fotografica dell'epigrafe non si trovino inseriti nell'immagine stessa: sarebbe infatti possibile (e ben più facile sul piano tecnico) realizzare immagini predefinite contenenti magari il calco sovrapposto alla fotografia o delle etichette di testo, ma questo risulterebbe

contrario allo sforzo di mantenere il contenuto quanto più separato possibile da una qualsiasi forma specifica: una volta preparate queste immagini, il composto di fotografia più calco o etichette non sarebbe più scomponibile, e sarebbe quindi presentabile all'utente finale nel solo modo previsto dall'editore, proprio come una fotografia stampata su un volume cartaceo. La struttura qui realizzata invece prevede di non alterare in alcun modo le immagini originali, ma semplicemente di affiancare loro dei file autonomi contenenti tutta la meta-informazione relativa a mappe, strati, didascalie e misurazioni; sarà poi compito del *software* unire questi due tipi di informazione (le immagini da un lato, le informazioni ad esse relative dall'altro) presentandole all'utente come un'unità in realtà solo apparente dal punto di vista della loro collocazione nella banca dati. Appare qui particolarmente evidente come l'informazione sia sempre mantenuta distinta nella sua specificità (in tal caso, informazione grafica e informazione meta-grafica), e del tutto indipendente da una particolare veste editoriale, dato che l'informazione meta-grafica rappresenta un mero testo XML che sarà trasformato in un aspetto visibile sulla fotografia originale dal programma di consultazione: un testo si fa dunque immagine, all'atto stesso della fruizione dell'edizione da parte dell'utente finale, e più elementi distinti vengono composti a formare un insieme complesso. Il ruolo della trasformazione e la scomposizione in elementi atomici che ne è alla base, si da risultare del tutto indipendenti da una qualsiasi forma predefinita, non potrebbe essere più evidente. D'altronde, aspetti particolarmente interattivi come mappe e misurazioni virtuali non avrebbero possibilità di essere pubblicati in altro modo diverso da quello dinamico: una mappa ha senso nel consentire l'individuazione di testo nella fotografia o viceversa, e le misurazioni vengono addirittura eseguite (sia pure in modo virtuale) dall'utente stesso dell'edizione.

Misurazioni paleografiche virtuali e materiale multimediale

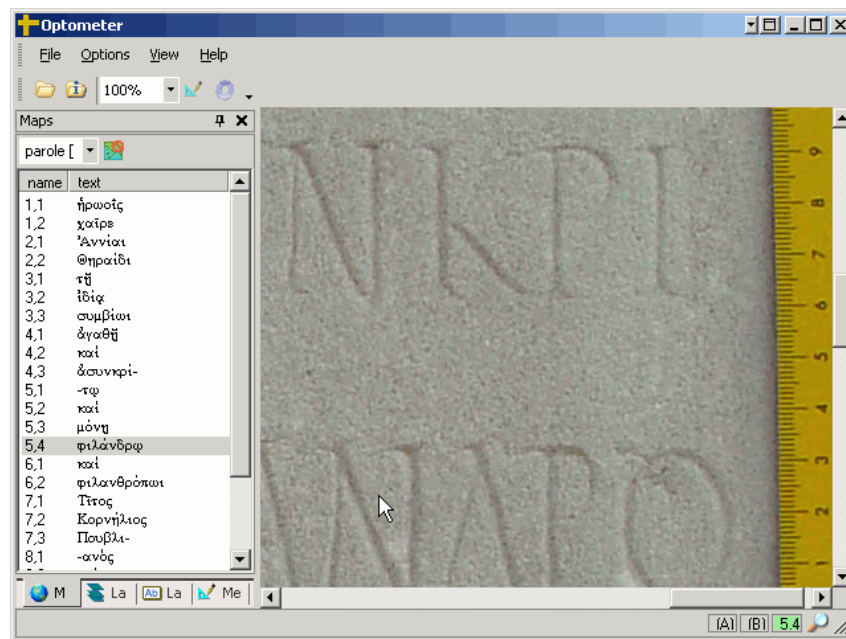
Uno dei vantaggi della natura multimediale e dinamica dell'edizione epigrafica elettronica, facilmente esemplificabile e relativo sia all'editore che all'utente finale, è inoltre costituito da un banale aspetto paleografico: è infatti buona prassi editoriale corredare ogni epigrafe di almeno alcune misurazioni ritenute più essenziali per una migliore descrizione paleografica, come ad esempio l'altezza media delle lettere, la larghezza media del loro solco, etc. Naturalmente ovvi motivi pratici inducono qualsiasi editore a limitare le misurazioni agli aspetti ritenuti a vario titolo più significativi, trascurandone altri che magari risulterebbero eccessivi per un'opera non specialistica, o che comunque anche in quest'ultima si potrebbero ritenere meno essenziali e dunque trascurabili. Anche un'edizione con specifici interessi paleografici infatti non potrebbe che offrire sempre un insieme finito rispetto alle innumerevoli misurazioni che potrebbero essere fatte sulla pietra per i motivi più disparati da parte degli utenti dell'edizione, e che naturalmente un editore non può né deve poter prevedere a priori: ad esempio, oltre alla semplice altezza media del rigo un utente interessato a una specifica epigrafe potrebbe ritenere necessari ai propri fini misure relative all'altezza di ogni singola riga e delle varie interlinee, o relative alla lettera più grande e alla più piccola, o all'estensione delle apicature in una lettera, o alla differente dimensione di lettere tondeggianti e lettere composte di sole aste, etc. È insomma facile osservare che nessun insieme di misurazioni fatte a priori potrebbe pretendere di soddisfare le esigenze più disparate, anche se ovviamente per lo più la scelta degli editori è orientata ai dati più significativi e quindi probabilmente in grado di appagare pienamente le necessità della maggior parte dei lettori.

In un'edizione elettronica diviene comunque possibile un approccio diverso, anche se in certa misura complementare: oltre a riportare un predefinito insieme di misurazioni fatte a priori dall'editore, essa può consentire allo stesso utente di effettuare un infinito numero di altre a seconda dei suoi specifici interessi. Affiancando infatti al testo una o più immagini fotografiche dell'epigrafe ripresa accanto a una scala metrica si ha la possibilità di utilizzare l'immagine di quest'ultima per dedurre in modo approssimativo la lunghezza di un qualsiasi segmento tracciato dall'utente (ad esempio utilizzando un *mouse*) sull'immagine stessa. Si tratta dunque in fondo di avvalersi del tradizionale riferimento visivo della striscia metrica, utile per valutare con uno sguardo le dimensioni approssimative dell'oggetto, per attingere però a una precisione assai maggiore di quella raggiungibile dagli occhi del lettore, dato che il programma di consultazione può misurare con precisione ogni lunghezza in rapporto a quella della striscia metrica e operare istantaneamente le dovute proporzioni (tenendo conto anche del fattore di ingrandimento o riduzione nella visualizzazione) per riportare la misurazione corrispondente. In un simile programma, l'utente non dovrà ad esempio che cliccare su due punti qualsiasi dell'immagine fotografica per ottenere immediatamente la misura della loro distanza. Naturalmente, si deve tener presente che simili misurazioni sono in vario grado approssimative, dipendendo da fattori come la risoluzione della fotografia ed eventuali distorsioni ottiche, ma in generale esse risultano assai precise in relazione agli scopi con cui sono effettuate, tanto più che il loro fine principale non consiste tanto nel produrre una cifra assoluta quanto piuttosto nell'offrire un termine di paragone rispetto ad altri tracciati nella stessa o in altre epigrafi. Peraltro, il programma che consente tali misurazioni è in grado di fornire

una misura approssimativa dell'errore statistico ad esse connesso, in dipendenza dalla risoluzione dell'immagine fotografica: è infatti ovvio che un'immagine dove 50 punti sullo schermo corrispondano a 1 centimetro potrà fornire delle misurazioni intrinsecamente più precise di una dove a 1 centimetro corrispondano solo 10 punti: in entrambi i casi comunque il programma offre sempre la stima di tale errore sì da consentire una maggiore precisione nella raccolta dei dati⁶.

Questo tipo di interattività offre dunque un ulteriore esempio di come sia possibile sfruttare a fondo le specificità di una 'vera' edizione elettronica, e insieme illustra l'intrinseca dinamicità della sua pubblicazione, cui in tal caso concorre lo stesso utente finale: a lui infatti viene affidato il compito di misurare un qualsiasi aspetto dell'epigrafe secondo i propri interessi, disponendo solo di un'immagine fotografica elettronica e di un opportuno programma di consultazione.

Trattandosi di un progetto in corso di realizzazione, è possibile qui illustrare più concretamente questi aspetti mostrando qualche immagine del *software* realizzato sinora allo scopo di visualizzare e modificare l'informazione meta-grafica connessa alle fotografie delle epigrafi:

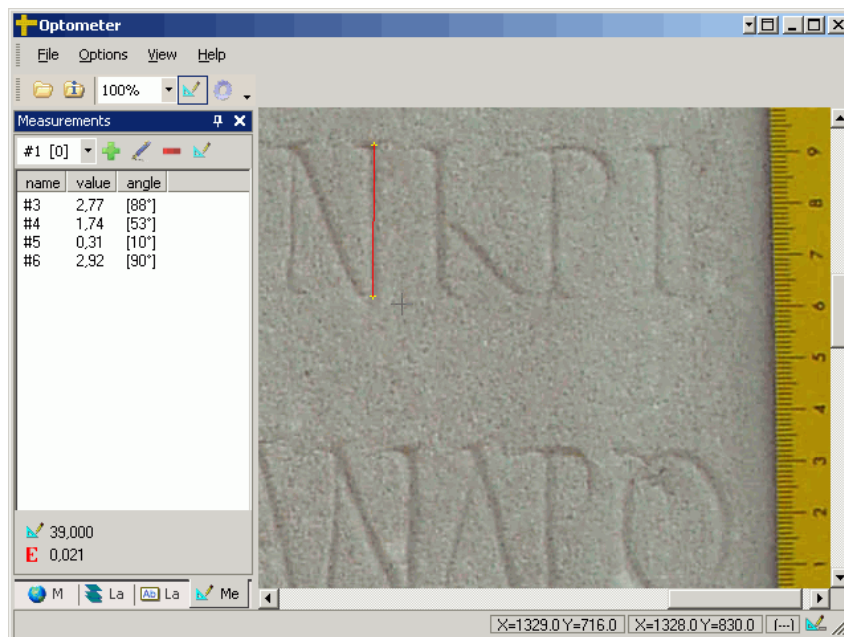


mappe dell'immagine

Nella figura qui sopra si mostra il programma deputato alle misurazioni paleografiche virtuali e alla visualizzazione di tutti gli altri elementi associabili a un'immagine (mappe, strati, didascalie): nel caso della figura il pannello di sinistra riporta il testo greco corrispondente alle varie regioni definite dalla mappa della fotografia (una regione per ogni parola): si può osservare come spostando il cursore sull'immagine il pannello di sinistra si sincronizzi con il testo ad essa corrispondente (φιλάνθρω).

Di seguito invece si riporta un esempio di misurazione virtuale:

⁶ Questo è uno dei motivi per cui di norma l'edizione offrirà più immagini fotografiche di una stessa epigrafe, non solo allo scopo di inquadrarne il contesto e documentarne tutti gli aspetti materiali, ma anche per fornire ove opportuno immagini di particolari ritenuti interessanti (a parità di risoluzione della fotocamera, la fotografia di una superficie di minore estensione fornirà ovviamente un'immagine dove più punti corrispondono al medesimo tratto).

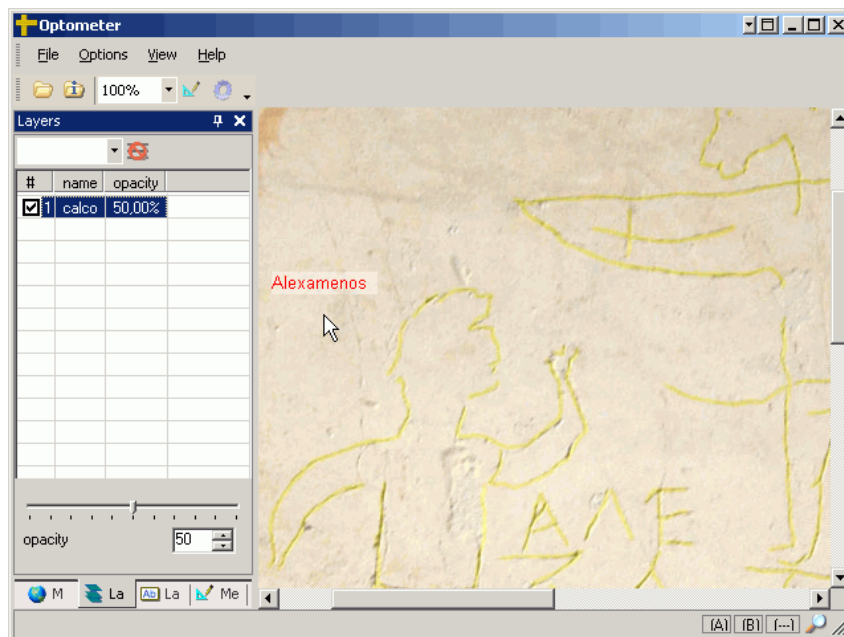


misurazioni virtuali

Questa volta sulla sinistra è visualizzato il pannello delle misurazioni, e in rosso appare la linea tracciata col mouse dall'utente per misurare l'altezza della lettera N. Come si vede, per ogni misurazione viene riportato il valore (in tal caso in centimetri) e l'angolo formato dalla misura rispetto all'asse orizzontale (90 gradi nel caso della misura relativa alla N). È inoltre possibile associare a ogni misurazione un nome specifico (ad es. "asse della N") e qualsiasi nota aggiuntiva di commento; una volta terminate le misurazioni, è possibile esportarle in un file (tipicamente XML) assieme a tutta questa informazione aggiuntiva e alle stime di errore, sì da poter utilizzare questi dati per qualsiasi scopo in applicativi di terze parti (ad esempio in un foglio elettronico come *Microsoft Excel*, che come gli altri programmi della suite *Office* è in grado di importare dati XML). Nella porzione inferiore del pannello delle misurazioni appare il valore dell'unità di misura (39,000, ovvero 1 cm = 39.0 punti) e la stima statistica dell'errore calcolato in base alla risoluzione dell'immagine fotografica (qui 0,021, sicché ad esempio 2,92 cm – l'altezza della N – va intesa con un margine di errore di 0,021 cm, cioè come $2,92 \pm 0,021$).

Naturalmente come per ogni altra modalità operativa del programma è possibile visualizzare la fotografia a diversi livelli di ingrandimento, in modo da poter effettuare misurazioni con la massima precisione (il fattore di ingrandimento o riduzione viene ovviamente tenuto in considerazione nel calcolo della misura).

Infine, l'immagine che segue mostra insieme strati, calco digitale e didascalie applicati in questo caso a un graffito (il celebre graffito di Alexamenos del Palatino):



calco e didascalie sovrapposti alla fotografia

La fotografia originale del graffito si affianca al suo calco digitale (una miniatura di entrambi è riportata qui sotto per riferimento⁷), che viene inserito in un gruppo di strati associati all'immagine di base: nel caso della figura, il calco è visualizzato (in giallo) come strato con un'opacità pari al 50%, lasciando così trasparire sullo sfondo la fotografia di partenza.



fotografia



calco digitale

In tal caso il pannello di sinistra mostra gli strati sovrapponibili alla fotografia originale (qui il solo calco), e consente (in basso) di regolare il livello di opacità di ogni strato o di nascondere o visualizzarlo a piacimento (cfr. il segno di spunta accanto al calco). Inoltre, alla fotografia sono aggiunte anche delle didascalie accanto ai personaggi: nella figura si vede quella relativa ad Alexamenos, in rosso: per ogni didascalia è infatti possibile specificare colore, rotazione, dimensioni e tipo di carattere.

Si deve comunque osservare che tutta l'informazione visualizzata in questi esempi sotto forma grafica in realtà non solo è memorizzata in modo del tutto distinto dall'immagine fotografica originale, ma come si è già detto appare strutturata in un semplice testo XML, che viene interpretato e trasformato in

⁷ Naturalmente il calco qui riportato è assai approssimativo, servendo al solo scopo di mostrare un esempio più adatto all'uso di didascalie e insieme l'applicabilità degli stessi strumenti a materiale relativamente eterogeneo.

forma visuale in tempo reale da un apposito programma. In tal modo diviene possibile aggiungere a una singola immagine un qualsiasi numero e tipo di osservazioni diverse, senza alcuna modifica all'architettura dei dati e alle fotografie originali: per una medesima fotografia si potranno ad esempio aggiungere strati per calchi digitali o evidenziazione di particolari regioni dell'immagine, mappe per la localizzazione di tutte le aree per qualche aspetto (a iniziare dal testo epigrafico) notevoli, e didascalie collocate in corrispondenza di qualsiasi elemento degno di nota, siano questi lettere di un testo (magari a evidenziare aspetti paleografici o materiali o linguistici o contenutistici), figure di un graffito, disegni e ornamenti affiancati all'iscrizione, etc. Tutta questa meta-informazione può essere aggiunta all'edizione in qualsiasi momento, e servire a intenti anche assai diversi: ad esempio si potrebbe pensare di espandere l'edizione per accogliere meta-informazione relativa all'evidenziazione del testo corrispondente sulla pietra a scopo didattico per un pubblico più o meno specializzato (come i visitatori di un museo o di un sito *web*, o l'utente non specialista di un *cd-rom*, o gli studenti di un corso elementare di epigrafia), oppure all'opposto per specializzare in senso molto tecnico l'edizione aggiungendo osservazioni specifiche di natura magari paleografica o materiale. In ogni caso, anche nella gestione dell'informazione multimediale si realizza così il modello fortemente vocato all'espansione e alla trasformazione già illustrato per altri aspetti del progetto.

Ulteriori applicazioni grafiche: repertorio paleografico virtuale e modelli neurali

La disponibilità di fedeli calchi digitali direttamente tratti dalle fotografie consente inoltre ulteriori interessanti applicazioni nell'ambito di un'edizione elettronica che si configura anzitutto come attivo strumento di ricerca. Specialmente nella prospettiva di una potenziale scalabilità dell'edizione infatti appare evidente la possibilità di estrarre dai calchi le sagome delle singole lettere, per poi costituire dei veri e propri repertori paleografici virtuali, contenenti la forma di ogni lettera dell'alfabeto così come appare sulla pietra dei testi presenti nel *corpus*.



calco digitale

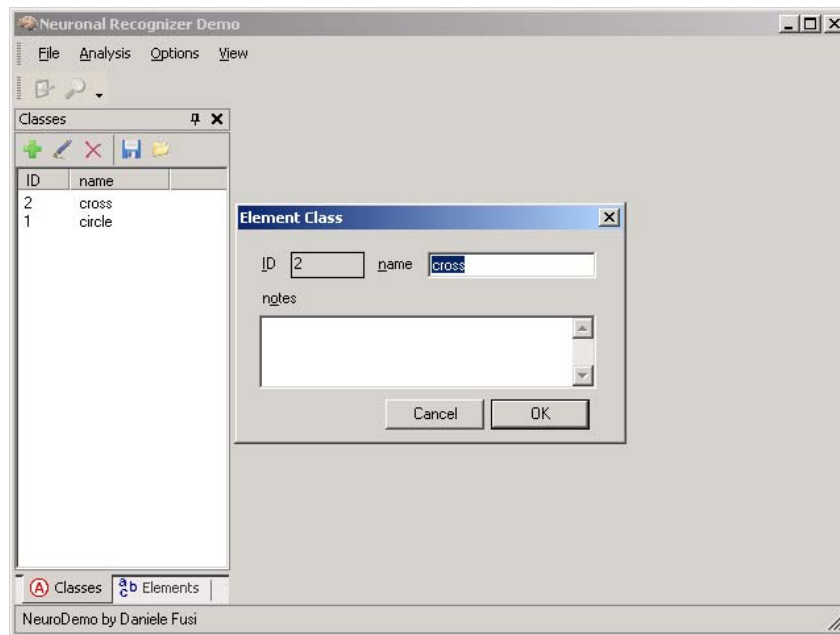


alcune lettere estratte dal calco

Questa operazione può risultare ancor più utile laddove il *corpus* risulti sufficientemente ampio e cronologicamente esteso, nella misura in cui diviene possibile assegnare la specificità di alcuni tratti delle lettere a determinate epoche sulla base di criteri non paleografici: in tal modo si definiscono (per quanto approssimativi) dei punti di riferimento lungo l'asse temporale dell'evoluzione del tratto grafico, rispetto ai quali diviene possibile fornire degli indizi di datazione su base paleografica per altre epigrafi del *corpus*. Tale repertorio diviene dunque uno strumento prezioso non solo per l'aspetto strettamente paleografico, ma anche per la sua capacità di offrire indizi cronologici sulla base della semplice analisi della forma delle lettere di un testo epigrafico. L'edizione elettronica consente peraltro di estendere questo tipo di analisi comparativa su una scala difficilmente gestibile con strumenti tradizionali, affidando a programmi specializzati il compito di effettuare i confronti anche su migliaia o decine di migliaia di campioni attinti dal repertorio interno al *corpus*. A questo scopo si possono utilizzare con profitto tecniche di analisi neuronale che rendono il programma capace di apprendere e generalizzare a partire da un insieme di campioni: nella fattispecie si tratta dunque di presentare al programma un numero sufficiente di esempi di forme per ciascuna lettera all'interno di un lasso cronologico predefinito, traendole direttamente dai calchi digitali, perché esso sia in grado di formarsi un modello paradigmatico per ogni forma, rispetto al quale valutare ogni nuova lettera incontrata. Ad esempio, dopo aver addestrato il programma a riconoscere un alfabeto definito come caratteristico di

I-II secolo d.C. semplicemente presentandogli varie decine di esempi per ogni lettera, dinanzi a una lettera appartenente a un'epigrafe del tutto estranea al *corpus* e dunque mai incontrata in precedenza il programma sarà in grado di giudicare il suo grado di somiglianza con l'idea che esso si è formato per ogni lettera degli alfabeti del I-II secolo, e di fornire quindi un preciso indizio cronologico per una datazione su base paleografica.

Per mostrare in termini estremamente semplici le potenzialità di un simile approccio e insieme confermare ancora una volta lo stato di progetto in concreto corso di realizzazione è possibile illustrare qui un breve esempio dimostrativo, rappresentato da un apposito programma. Quest'ultimo presenta all'utente una semplice interfaccia grafica con la quale è possibile definire in modo del tutto arbitrario delle classi di elementi, addestrare il programma a riconoscere ciascuna di essa e verificare immediatamente la sua capacità di apprendimento. Si consideri dunque per semplicità un esempio banale come un ipotetico linguaggio comprendente due soli segni, rappresentati graficamente da un pallino (o) e da una croce (x), disegnati con tratti anche molto diversi fra loro. Anzitutto il programma viene istruito a considerare l'esistenza di questi due segni, definendo due classi distinte, pallino e croce, come mostra la seguente schermata:

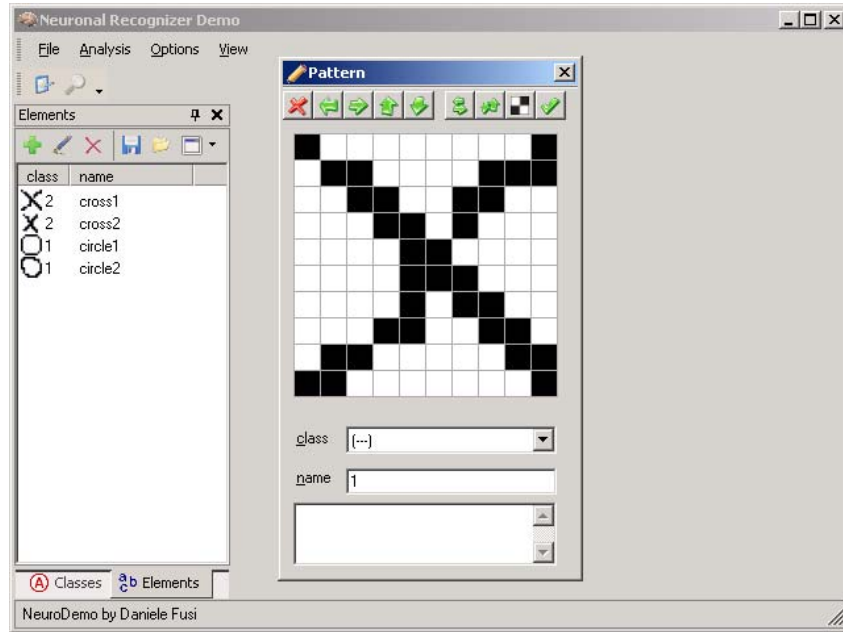


dimostrazione di apprendimento neurale: definizione delle classi

Come si vede, l'utente definisce due classi semplicemente dando loro dei nomi (qui *circle* e *cross*) ma senza fornire alcuna informazione relativa alla loro forma. Dunque il programma sa solo che nell'ipotetico linguaggio che dovrà interpretare sono presenti due classi di segni (cfr. il pannello di sinistra nella finestra dell'interfaccia riportata sopra), ma non ha il minimo dato su quale sia la forma di ciascuno di essi. Questo tipo di informazione non costituirà un dato aprioristico, ma dovrà essere appresa dal programma stesso per via di analisi e confronto di tutti gli esempi dei segni che gli saranno presentati, esattamente come accade per l'apprendimento nel cervello umano.

L'utente dunque fornisce ora una serie di figure, semplicemente disegnandole su una griglia⁸ sullo schermo per mezzo di un *mouse* e assegnandole di volta in volta a una delle due classi (croce o pallino): in tal modo presenta al programma solo esempi concreti, lasciando ad esso il compito di astrarre e formarsi da essi un'idea della forma paradigmatica di ciascuna classe di segni.

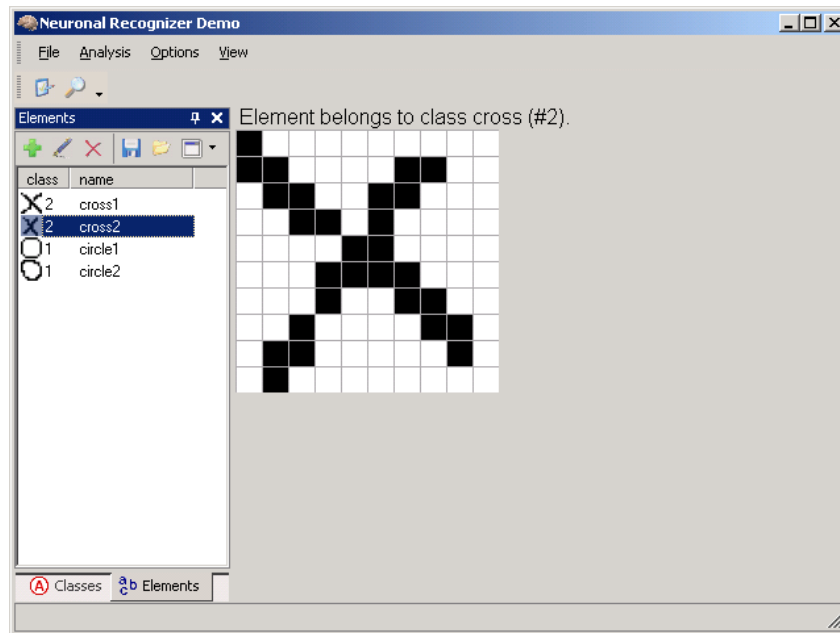
⁸ La griglia dell'esempio non è che la rappresentazione ingrandita della digitalizzazione di un'immagine, cioè della sua descrizione per mezzo di punti discreti, in tal caso semplicemente neri o bianchi e limitatamente a un quadrato di 10 x 10 punti. Naturalmente nel caso di una lettera tratta da un calco digitale i punti che la definiscono saranno migliaia, e potranno anche essere dotati di intensità cromatiche diverse (su una scala di grigi dal nero al bianco); peraltro, tali lettere saranno comunque sottoposte a un'adeguata preparazione prima di essere presentate al programma per l'apprendimento (ad esempio normalizzando le dimensioni delle lettere e convertendo la loro figura in una serie di contorni, che sono i tratti percepiti dal cervello come pertinenti per il riconoscimento di una sagoma).



dimostrazione di apprendimento neurale: tracciamento di un esempio per la classe X (nel pannello di sinistra si possono osservare le miniature di altri campioni appartenenti alle due classi X e O)

Come si vede nella figura riportata sopra, l'utente ha qui presentato al programma in tutto quattro esempi concreti di croci e pallini, tracciati in modo diverso e senza alcuna cura di precisione sulla griglia offerta dall'interfaccia; ognuno di questi disegni è stato associato dall'utente a una delle due classi. Una volta presentati gli esempi, viene chiesto al programma di analizzarli per apprendere da essi un'idea delle classi di questo ipotetico linguaggio binario. Si tratta naturalmente di un procedimento da effettuarsi una volta per tutte, almeno finché non si presentino nuovi dati per l'apprendimento, sicché il programma offre la possibilità di salvare i risultati della sua elaborazione per poterli poi recuperare ogni volta che serve.

Compiuto infine l'apprendimento sulla sola base degli esempi concreti di pallini e croci disegnati dall'utente il programma è quindi pronto per tentare di riconoscere qualsiasi tracciato e assegnarlo a una delle due classi: l'utente può ora disegnare un segno che il programma non ha mai incontrato prima, ma che esso può comunque provare a giudicare come più o meno simile a una delle classi di segni di cui si compone l'ipotetico linguaggio che ha imparato a riconoscere. È dunque sufficiente che l'utente tracci un disegno con il *mouse* perché il programma tenti di riconoscerlo in base a quanto ha appreso in precedenza:



dimostrazione di apprendimento neurale: il programma riconosce come appartenente alla classe X una forma mai incontrata prima sulla base dell'apprendimento effettuato su forme simili presentatigli in precedenza

Nell'esempio riportato qui sopra l'utente ha appena tracciato il disegno di una croce in un modo mai fatto prima per nessuno dei campioni presentati al programma durante l'apprendimento (cfr. ad esempio il tracciato della croce riportato nella figura qui sopra e quello della figura immediatamente precedente), e ha chiesto al programma di tentare di riconoscere a quale classe appartenga, ovvero se sia un pallino o una croce: come si vede in tal caso si tratta di una croce, correttamente riconosciuta. Pertanto, semplicemente dichiarando l'esistenza di due tipi di segni e fornendo un paio di esempi per ogni tipo l'utente ha reso il programma in grado di apprendere dagli esempi presentati, e utilizzare le nozioni apprese per classificare ogni nuovo disegno ad esso proposto⁹.

Naturalmente questo è solo un esempio estremamente semplificato, ma è facile comprendere come un simile programma possa riconoscere le diverse forme delle lettere estratte dai calchi digitali di un *corpus* epigrafico elettronico: in questo caso le classi definite aprioristicamente non saranno più le due dell'esempio (croce e pallino), ma varie centinaia corrispondenti a più tipi di alfabeto, ciascuno collocato cronologicamente sulla base di criteri non paleografici relativi alle iscrizioni che lo attestano (tale forma di *alpha* per il IV secolo a.C., tale forma per il I a.C., tale per il III d.C., etc.). Il programma potrà quindi effettuare il proprio addestramento su questi modelli, e poi rispondere alle richieste di riconoscimento di forme di lettere mai incontrate prima, sì da giudicarle più o meno simili all'uno o all'altro tipo di alfabeto e quindi più o meno vicine all'uno o all'altro periodo cronologico. Si realizza in tal modo un vero repertorio paleografico interattivo, capace di fornire uno strumento di ricerca tanto più potente quanto più vasto sarà il *corpus* da esso dominato e di apprendere in modo dinamico dal materiale che viene presentato e aggiunto di volta in volta.

b) Aspetto testuale: dal testo marcato alla banca dati

Se multimedialità, interattività e applicazioni grafiche più o meno evolute risultano abbastanza intuitive dalle specificità dell'edizione propriamente digitale, forse meno immediate ma più rilevanti sono poi le conseguenze di tale specificità nella strutturazione dell'informazione testuale. Come si è accennato a proposito dell'esempio di un vocabolario digitale, la mera trasposizione di un testo su un *medium* diverso non può che superficialmente considerarsi una vera edizione digitale: perché quest'ultima sia fruibile in modi e vesti virtualmente infiniti è necessario anzitutto che il suo contenuto sia adeguatamente strutturato dal punto di vista semantico. Se ci si limita al semplice testo, i livelli di semantizzazione che esso può assumere nella sua forma digitale possono essere di grado molto vario: uno minimo può essere ad esempio rappresentato dalla semplice trasposizione di un testo continuo da

⁹ Si noti peraltro che la natura neurale di questo programma consente ad esso anche di non dare affatto una risposta, qualora i campioni usati per l'apprendimento non risultino sufficienti per il riconoscimento; in tal caso, invece di produrre un (rischioso) falso positivo il programma risponderà semplicemente che non è in grado di riconoscere con sufficiente certezza una data forma.

una pagina di carta a un testo elettronico, riportandolo magari su un applicativo di videoscrittura. È evidente che in tal caso le possibilità di fruizione di un simile testo saranno assai limitate, non andando molto oltre la lettura a video o la ricerca a tutto testo di una parola al suo interno. Un'operazione molto più fruttuosa può invece essere rappresentata dall'attribuzione di una qualche forma di marcatura semantica a questo testo: ad esempio, in un testo come:

A • FABIO • A • F • FABI-
ANO • EQVITI
ROMANO • DECV-
D RIONI • OST M
 QVI • VIXIT • ANNIS •
 XXII • M • XI • D • XX •
A • FABIVS • TROPHIMVS
 PATER • FECIT

potrebbero essere in qualche modo marcate le abbreviazioni (*A(ulus)*, *A(uli)*, *F(abi)*, etc.), le cariche del personaggio (*equiti Romano*, *decurioni*), le lettere indicanti numeri (*XXII*, *XI*, *XX*), e quant'altro si desidera individuare come rilevante o disambiguare nel flusso altrimenti indistinto di testo. In effetti, molte edizioni digitali recenti adottano questo tipo di strategia, partendo dal testo di base per applicarvi marcature più o meno raffinate e complesse, solitamente implementate utilizzando XML. Simili edizioni hanno indubbiamente numerosi vantaggi, e possono risultare più che soddisfacenti rispetto agli scopi proposti dai loro autori: esse realizzano infatti un certo livello di semantizzazione del dato testuale, che non è più un flusso indistinto di lettere ma viene opportunamente caratterizzato secondo il tipo di informazioni di cui è portatore; tecnologie come XML e Unicode, ampiamente utilizzate a questi scopi, risultano poi molto valide nella misura in cui rappresentano degli standard consolidati e universalmente diffusi, e danno luogo a un modello di edizione molto compatta e semplice, dove tutta l'informazione è contenuta in un file di testo marcato, interpretabile e fungibile a diversi livelli e facilmente utilizzabile anche come piattaforma di scambio in virtù della intrinseca vocazione alla trasformazione propria di XML (anzitutto tramite tecnologie complementari e altrettanto diffuse come XSLT). Un semplice applicativo di videoscrittura sul versante redazionale e un *browser web* da parte del fruitore dell'edizione sono in tal caso già strumenti sufficienti per la realizzazione di edizioni non troppo complesse e abbastanza versatili, benché richiedano all'editore lo sforzo di digitare testo seguendo un determinato schema di marcature definito a priori e conforme allo standard XML.

Tuttavia almeno rispetto agli scopi prefissati per un progetto come quello qui presentato un simile modello editoriale comporta anche qualche decisivo svantaggio, connesso al fatto che come per il semplice testo cartaceo riportato su *medium* digitale anche in tal caso è appunto un testo continuo, per quanto accuratamente marcato, che funge da struttura portante dell'intera edizione. Ciò rappresenta un limite non solo in rapporto alle specificità del testo epigrafico, dove il complesso di tutta l'informazione non testuale associata per vari motivi (archeologici, artistici, paleografici, filologici, storici, linguistici, etc.) all'iscrizione finisce spesso per essere quantitativamente molto superiore e notevolmente varia rispetto al testo in sé, ma anche rispetto alle intrinseche limitazioni delle tecnologie di marcatura. Per fare un esempio banale, si consideri di nuovo il testo citato sopra; in un simile testo si potrebbero voler marcare le abbreviazioni in quanto tali con il relativo scioglimento (riquadro 1), e magari anche i vari componenti del nome dei personaggi con le relative cariche e ruoli (ad es. a fini prosopografici; riquadro 2): in tal modo si dovrebbe ad esempio marcare in qualche modo la prima *A* come abbreviazione di *A(ulo)*, ma anche come prenome del personaggio, allo stesso modo in cui si marcherebbero *Fabio* e *Fabiano* come *nomen* e *cognomen*; contemporaneamente, il patronimico *A(uli) F(ilio)* dovrebbe essere marcato sia in quanto tale che come duplice abbreviazione, nonché rientrare nella più estesa marcatura relativa al nome del personaggio nel suo insieme (dal prenome sino al *cognomen*), etc. Già per queste poche informazioni metatestuali si dovrebbero applicare marcature a una buona porzione del testo e in qualche modo sovrapposte le une alle altre.

D • M • A • FABIO • A • F • FABI-
ANO • EQVITI
ROMANO • DECV-
RIONI • OST
QVI • VIXIT • ANNIS •
XXII • M • XI • D • XX •
A • FABIVS • TROPHIMVS
PATER • FECIT

D • M • A • FABIO • A • F • FABI-
ANO • EQVITI
ROMANO • DECV-
RIONI • OST
QVI • VIXIT • ANNIS •
XXII • M • XI • D • XX •
A • FABIVS • TROPHIMVS
PATER • FECIT

Qualsiasi aggiunta di dati relativi poi a interessi specifici, siano essi di natura epigrafica, paleografica, archeologica, artistica, storica, filologica, linguistica etc., dovrebbe poi essere in qualche modo inserita nel sistema delle marcature, che finirebbe così per crescere su se stesso generando un'enorme nidificazione e complessità, di ostacolo sia all'inserimento dei dati che al loro trattamento automatico; inoltre per questa via si giungerebbe comunque al limite oggettivo imposto da una tecnologia di marcatura come XML, che non consente la sovrapposizione di diverse marcature, allo stesso modo in cui sarebbe impossibile evidenziare con diversi colori la stessa porzione di testo. Appare dunque chiaro che un'edizione basata esclusivamente sul testo marcato come struttura portante finisce per costringere il dato metatestuale, la cui rilevanza per documenti epigrafici è spesso quantitativamente maggiore del testo stesso, e che in ogni caso deve poter essere concepito come espandibile in qualsiasi momento senza pregiudizio per il materiale esistente e la sua struttura.

Il modello qui proposto pertanto, dovendo adeguarsi agli obiettivi evidenziati sopra, abbandona l'impostazione più strettamente 'tipografica' dell'edizione basata sul testo marcato, operando in una prospettiva più latamente e astrattamente 'informatica': non si tratta più di partire da un testo ideale o concretamente preesistente da marcare con opportuni *tag*, ma piuttosto di definire una struttura dei dati più remota da una qualsiasi specifica incarnazione formale, sia essa propria della tipica scheda epigrafica, di un modello editoriale cartaceo o di altri tipi di presentazione. Naturalmente la scheda epigrafica rimane l'ideale punto di partenza per la definizione del tipo di materiale da trattare, ma la sua organizzazione interna non assume necessariamente la scheda come modello di strutturazione dei dati. Di fatto, quanto giace 'dietro le quinte' dell'edizione (o meglio delle edizioni potenzialmente pubblicabili a partire dall'unico insieme di dati) ha una struttura definita anzitutto dalla necessità da un lato di scomporre l'informazione fino al suo atomo, e dall'altro di mantenere i contenuti quanto più possibile indipendenti da una qualsiasi forma specifica. Non si tratta più di assumere alla base un testo epigrafico per poi inserirvi per quanto possibile ogni altra informazione tramite marcature più o meno complesse, ma di costituire una vera e propria banca dati epigrafica dove ogni tipo di dato, incluso il testo, viene scomposto nei suoi elementi e memorizzato a parte secondo le sue specificità; a partire da questa banca dati poi apposito *software* provvederà a recuperare i vari elementi, a ricombinarli e presentarli nel modo e nella veste più adatti alla richiesta dell'utente. In questo modello dunque si può assimilare l'edizione non più a un testo sequenziale marcato, ma a una sorta di albero dell'informazione dal cui tronco si dipartono molteplici diramazioni, ciascuna capace di ospitare un determinato tipo di dato e di crescere virtualmente all'infinito sulla stessa pianta, con nuovi rami che rampollano dal tronco esistente la cui struttura e contenuti rimangono nondimeno inalterati. Fuor di metafora, questa struttura viene nondimeno realizzata con le medesime tecnologie impiegate da altri tipi di edizione: come vi sono dei testi marcati rappresentati da file XML che costituiscono tutta l'opera, così qui questi alberi di dati sono costituiti da semplici file XML che racchiudono tutti i dati dell'edizione in una struttura gerarchica. Mentre però nelle edizioni di altro tipo XML è usato come tecnologia per marcare il testo delle iscrizioni disposto in sequenza per formare il *corpus*, qui lo stesso XML è usato piuttosto come articolato contenitore di informazione, dove il testo è solo uno dei tanti contenuti e non rappresenta la base portante dell'opera. In entrambi i casi tutto il materiale dell'edizione è incluso in semplici file di testo (sia esso codificato in *Unicode* o in altri standard) strutturati secondo lo standard XML, producendo così una struttura estremamente compatta, portatile e trasformabile in altri formati. Tuttavia, nel caso di questo progetto XML non è solo un modo di inserire nel testo epigrafico dati non testuali, ma un astratto contenitore di informazione strutturata secondo i propri principi. In questo ambito una tecnologia come XML, oltre a rappresentare oggi lo standard di fatto nella codifica dei dati per trasmissione o scambio sulla rete, consente di dotare di struttura dati dalla complessità teoricamente illimitata, in uno schema ad albero dove ogni elemento di informazione può contenere un qualsiasi numero di altri elementi logicamente correlati: la scelta di questa tecnologia riguarda in primo luogo la necessità della sua massima apertura all'espansione e scambio dei dati, e insieme la relativa provvisorietà della struttura che si intende definire per l'informazione epigrafica nel suo complesso. Nell'ambito infatti di una fioritura sempre maggiore di numerosi e diversi progetti in questo settore, ciascuno dei quali frutto di esperienze, competenze e investimenti di tempo e risorse magari molto differenti ma comunque in grado di fornire apporti preziosi alla definizione di alcune basi comuni, il primo imperativo per le nuove iniziative dovrebbe essere costituito dalla capacità di apertura all'espansione sia in senso quantitativo che qualitativo e alla facilità di scambiare i dati incorporati da ogni progetto. In questo senso XML può fornire una base di partenza solida in virtù della diffusione di tale tecnologia e degli strumenti per gestirla, e insieme massimamente vocata alla trasformazione e alla duttilità, doti fondamentali specie per le fasi preliminari di ogni progetto che si muova su principi cui la mancanza di

standard conferisce ancora un notevole grado di labilità e provvisorietà, per quanta cura si possa impiegare nel concepirli e confrontarli con gli apporti di altri progetti più o meno simili.

Nel progetto qui proposto dunque la parte preponderante assegnata all'illustrazione dell'XML deriva anzitutto dall'osservazione di tali necessità, ed è funzionale allo sforzo di una definizione almeno minimale di una base comune di scambio di dati e della loro struttura; questo tuttavia non esclude affatto che nei dettagli dell'implementazione concreta si possa (e anzi debba) far posto anche ad altre tecnologie, in primo luogo al *database* relazionale rispetto al quale del resto XML non si è mai configurato come un antagonista, quanto piuttosto come un utilissimo complemento. In effetti, il progetto qui illustrato si avvale appunto di una base specificamente vocata alla trasformazione dei dati ai fini della loro importazione ed esportazione da e verso altri formati, siano essi standardizzati o proprietari, ma da tale base ricava automaticamente un nucleo dal contenuto più rigidamente strutturato destinato all'uso interno da parte dei programmi di consultazione e corrispondente a un *database* relazionale. In tal modo si possono coniugare l'ineguagliabile efficienza di quest'ultimo nella gestione e ricerca dei dati con la straordinaria versatilità e capacità di trasformazione proprie di XML: in un certo senso, XML rappresenta in questo ambito il volto esteriore e l'impianto di base del progetto, con il quale è possibile stabilire dialoghi con innumerevoli altri formati e *corpora*, ma per quanto riguarda la sua implementazione concreta esso nutre anche un nucleo interno dalla strutturazione più rigida ma proprio per questo più efficiente, rappresentato da un *database* relazionale automaticamente generato e mantenuto da un apposito *software* che funziona da ponte fra i due componenti. In questo senso la novità dei progetti basati in vario grado su XML non deve essere travisata come il frutto della contrapposizione fra tecnologie, ma piuttosto trovare la sua prima giustificazione proprio nelle notevoli possibilità di dialogo e scambio di dati e metadati che una simile tecnologia rende possibili e soprattutto agevoli, al di là delle specificità, per quanto numerose e idiosincratice, che ogni progetto possa definire al suo interno¹⁰.

Nell'ambito di questo progetto dunque XML rappresenta un ideale contenitore di informazione perché oltre alla sua natura standardizzata e alla sua ampia diffusione in altri tipi di progetti editoriali consente di raccogliere grandi repertori di dati, conferendo loro una ben definita struttura semantica ma lasciando contemporaneamente spazio a notevoli capacità di espansione e specializzazione senza alterare il materiale preesistente. In tale prospettiva XML, come uno degli standard per eccellenza nello scambio di dati fra sistemi diversi, rappresenta non solo un versatile intermediario per consentire lo scambio di informazioni tra diversi progetti editoriali, ma anche all'interno di questo singolo progetto un intermediario tra le diverse fasi redazionali dell'opera. La storia stessa di questo progetto infatti ha comportato e comporta tuttora una notevole fluidità nella definizione delle sue strutture e dei suoi contenuti, in funzione delle diverse finalità proposte: da un primo nucleo essenzialmente incentrato sul testo epigrafico e funzionale a precisi interessi linguistici, metrici¹¹ o comunque specialistici a una vera e propria edizione epigrafica di testi greci e latini classici, per poi estendersi da un lato al mondo medievale e dall'altro alla raccolta di una grande varietà di documenti non solo testuali relativi al teatro antico, il progetto si è evoluto ampliando la gamma dei suoi contenuti e specializzando in diverse direzioni i dati metatestuali, nonché seguendo l'evoluzione delle tecnologie informatiche degli ultimi anni. Questo percorso evolutivo ha dunque richiesto un continuo rimaneggiamento delle strutture concepite al suo avvio e del relativo *software* redazionale creato *ad hoc* per esse, implicando così anche molteplici trasformazioni dei contenuti già esistenti. In questo senso, il vantaggio di conservare tutta l'informazione in contenitori XML, capaci di memorizzare ogni dato al suo posto nella struttura ma nel contempo di espandersi e trasformarsi in modo assai semplice consiste così nel produrre una notevole longevità dell'insieme, suscettibile di rinnovarsi anche profondamente senza che i dati già esistenti vadano perduti. L'edizione digitale a monte non è che un insieme di *file* di testo, codificati secondo standard internazionali (in questo caso *Unicode*) e strutturati secondo la sintassi di tecnologie ormai universali come XML, cui eventualmente si collegano risorse multimediali (es. fotografie, calchi, filmati, etc.) che rimangono sempre autonome; ciò offre la migliore garanzia non solo di longevità per l'opera, ai cui contenuti è altrettanto facile accedere come a un qualsiasi semplice testo, ma anche di apertura all'espansione e alla trasformazione. A valle di questo insieme si trovano appunto in primo luogo quelle trasformazioni operate dal *software* che consentono di pubblicare l'unico contenuto nelle molteplici forme per i diversi pubblici e *media* che si desiderano considerare, e in prospettiva tutte le

¹⁰ Peraltro occorre anche guardarsi da approcci eccessivamente superficiali cui potrebbe indurre una tecnologia come XML, specie se associata in modi non sempre troppo consapevoli ad altre tecnologie sorelle come HTML e a una visione troppo legata alla tradizionale edizione cartacea, privilegiando una particolare e aprioristica forma piuttosto che derivare molteplici forme da un unico contenuto, strutturato secondo le specificità di quanto è una vera banca dati ancor prima che un semplice testo continuo dotato di annotazioni più o meno articolate.

¹¹ Sulla specializzazione dell'edizione digitale in senso linguistico e metrico e sulle potenzialità dell'uso di sistemi esperti anche a scopi editoriali si veda più oltre.

trasformazioni che potranno essere necessarie per riconvertire i contenuti in tecnologie future quando quelle di oggi risulteranno ormai obsolete.

La prospettiva qui adottata dunque riflette il passaggio da un'edizione concepita come puro testo marcato a una derivante per trasformazione da una vera e propria banca dati articolata, dove il testo è solo uno degli elementi memorizzati. La banca dati alla base dell'edizione opera una scomposizione semantica dell'informazione ulteriormente avanzata rispetto alla semplice marcatura di un testo, conducendo alla generazione di strutture la cui complessità richiede necessariamente l'ausilio di strumenti *software* per l'immissione dei dati a livello redazionale come per la loro trasformazione ai fini della pubblicazione, o meglio delle (molteplici) pubblicazioni, da essa derivate. La pubblicazione, ad esempio in forma di tradizionale scheda, non è quindi il modello di partenza per la strutturazione *interna* dei dati, ma il punto di arrivo di una serie di trasformazioni più o meno complesse che operano sulla banca dati epigrafica, riassembrandone i vari componenti opportunamente selezionati, filtrati e riordinati in tempo reale per soddisfare nel modo più adatto le richieste dell'utente. I dati in sé rimangono invece il più possibile 'amorfi', strutturati e atomici per consentire l'adattamento a qualsiasi forma e la loro selezione a partire da ricerche incrociate. Con ciò si ottiene la massima separazione fra l'unico contenuto e le molteplici forme che esso può assumere in fase di pubblicazione, all'interno di una struttura slegata dal testo e dettata solo da criteri semantici, dove il dato metatestuale è libero di espandersi e strutturarsi secondo le sue specifiche esigenze; inoltre, anche la realizzazione di questa banca dati almeno nella sua forma redazionale, basata su XML, consente la massima apertura all'espansione e alla trasformazione dei contenuti sia in senso quantitativo che qualitativo.

Esempi di realizzazione

Il modello qui adottato per la struttura della banca dati si basa dunque essenzialmente su considerazioni semantiche e logiche, che dalla tradizionale scheda epigrafica traggono un primo alimento per la definizione delle proprie categorie; queste tuttavia dietro le quinte sono opportunamente ristrutturare e scomposte per consentire il migliore trattamento automatico dei dati. Il primo requisito è infatti costituito dalla completa trattabilità automatica dei dati, anche a scapito della loro immediata leggibilità da parte di un utente umano; ciò garantisce la possibilità di trasformarli facendo loro assumere qualunque veste, e di poter attingere ad essi secondo le modalità e gli intenti più disparati. D'altra parte, appositi strumenti *software* di uso redazionale e destinati all'utente finale con ricca interfaccia grafica sono deputati a schermare il redattore o il lettore dalla complessità dei dati, garantendo la facilità della loro immissione e modifica, come del loro recupero da parte del lettore. Benché come già osservato i dettagli relativi all'implementazione prevedano anche l'uso di veri e propri *database* relazionali (ad es. per i risultati di indicizzazioni automatiche si da consentire un più veloce recupero dell'informazione in sede di consultazione), tali programmi operano essenzialmente generando o manipolando codice XML che costituisce la base di partenza della banca dati epigrafica propriamente detta. Sebbene la scelta di XML possa qui apparire meno ovvia nella misura in cui non si tratta di marcare (cioè strutturare) il testo di una scheda epigrafica, ma di realizzare una banca dati del tutto indipendente da una forma specifica, essa è comunque facilmente comprensibile considerando che:

- XML è uno standard ampiamente diffuso per lo scambio di dati complessi assieme alle loro stesse definizioni, nella misura in cui si tratta di un formato autodescrittivo; le sue applicazioni sono sempre più estese, e stanno modificando profondamente il panorama dell'informatica, dalla creazione di pagine HTML per il *web* al rilevante ruolo in sistemi di ultima generazione per la creazione di *software* come il *DotNet framework* promosso da *Microsoft*.
- XML è portabile (in quanto semplice testo e dunque *platform-independent*), aperto ad espansione ed estremamente duttile in quanto alla base di innumerevoli tipi di trasformazione creati per esso: un insieme di *tag* specializzati per la trasformazione di XML in altro XML (da semplice filtro o riordinamento dei dati a profonde alterazioni della loro struttura) o virtualmente in qualsiasi altro formato (HTML, RTF, PDF o altro formato tipografico via XSL:FO, etc.), definiscono un vero linguaggio di trasformazione (XSLT), che pure in sé altro non è che un testo XML, e alle sue capacità di trasformazione si possono aggiungere estensioni personalizzate a seconda del compito da affrontare.
- direttamente proporzionale al successo di XML è il supporto degli attuali programmi e strumenti di sviluppo *software* a questa tecnologia: la maggior parte del *software* di nuova generazione per il trattamento di dati consente di gestire XML, a cominciare da applicativi della diffusione di *Office*.
- XML affianca alla strutturazione dei dati la possibilità di una loro rigorosa descrizione formale tramite la creazione di schemi, anch'essi rappresentati da semplici file XML creati avvalendosi di una speciale serie di *tag* definita appositamente per la descrizione semantica del contenuto e del tipo di ogni dato (XSDL). Questo da un lato rende l'XML completamente autodescrittivo, e dall'altro

consente di disporre costantemente di un riferimento preciso in ogni suo dettaglio per la strutturazione dei dati e la convalida della loro forma e contenuti, garantendo in tal modo l'integrità della banca dati.

- con la sua strutturazione 'ad albero' XML consente di descrivere strutture di alta complessità, ben adatte a materiale come testo, non facilmente costringibile in schemi troppo rigidi almeno nelle prime fasi della progettazione dell'edizione digitale, tanto più in funzione della massima apertura all'espansione e allo scambio dei dati, cui qualsiasi progetto in ambito epigrafico dovrebbe tendere.

Un semplice esempio che illustra quest'ultimo aspetto e insieme mostra in modo evidente il tipo di scomposizione cui è sottoposta l'informazione di una scheda epigrafica tradizionale può essere offerto dalla datazione di un'epigrafe, che deve poter coprire qualsiasi tipo di indicazione, dalla più generica (incluso il caso in cui la data sia semplicemente ignota) alla più precisa (con tanto di giorno e anno), e nondimeno deve poter essere espressa in modo da essere utilizzata automaticamente ai fini di ordinamento o filtro cronologico dei testi. Ad esempio, tra le varie forme in cui appare la datazione in una scheda epigrafica possono essere citate:

ignota
III secolo d.C.
prima del II secolo a.C.
dopo il I secolo d.C. (?)
fra II e III secolo d.C.
circa IV-III secolo a.C.
178 a.C.
136/5 a.C.
dopo il 178 a.C.
prima del 136/5 a.C.
150-125 a.C.
età di Tiberio (14-37 d.C.)
... etc. etc.

Dinanzi a una tale varietà appare evidente l'inadeguatezza di una semplice tabella a due dimensioni, che spesso viene mentalmente associata a insiemi di dati¹², quando non addirittura assunta a modello concreto per la loro strutturazione: per ogni datazione si dovrebbero infatti prevedere numerose celle in una ipotetica griglia, molte delle quali benché inevitabilmente presenti (dato che su ogni riga della griglia dovrebbe collocarsi una datazione diversa) rimarrebbero di fatto inutilizzate in quanto non pertinenti: ad esempio, si potrebbe pensare a una serie di celle come:

secolo	anno	a.C.	circa	forse	anno succ.	ignota	etc...
--------	------	------	-------	-------	------------	--------	--------

dove collocare i vari componenti della datazione, sicché una data come "III secolo d.C." potrebbe essere rappresentata come:

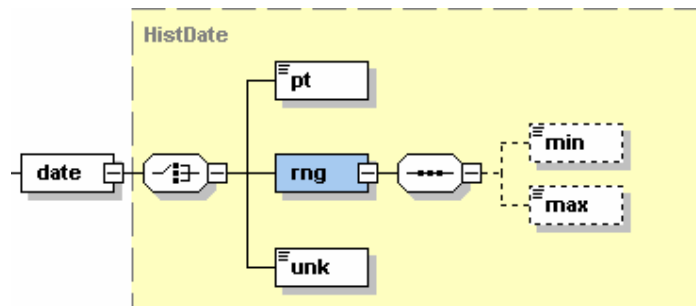
secolo	anno	a.C.	circa	forse	anno succ.	etc...
3	–	si	–	–	–	

una come 170 d.C. come:

secolo	anno	a.C.	circa	forse	anno succ.	etc...
2	170	no	–	–	–	

¹² Ancora una volta con ciò non si intende certo sostenere che XML rappresenti l'unico modo di implementare una simile strutturazione dei dati: sarebbe ad esempio possibile organizzare in modo concettualmente simile l'informazione anche in un *database* relazionale, ma quanto interessa far rilevare qui è anzitutto la necessità di un'adeguata scomposizione e strutturazione dei dati, trascurando i dettagli relativi all'implementazione concreta. Il confronto con un'astratta organizzazione superficialmente tabulare dei dati, che nulla ha (o dovrebbe avere) a che fare con la struttura di un *database* relazionale propriamente inteso, risulta comunque comodo per apprezzare la versatilità di XML nell'esprimere in modo insieme comprensibile e strutturato l'informazione qui esemplificata, tanto più che non pochi tentativi più o meno recenti di realizzare delle banche dati in ambiti umanistici si sono spesso limitati per motivi di ordine pratico o teorico a eccessive schematizzazioni dei dati in pesanti e inefficienti griglie monolitiche.

Anche questi semplici esempi mostrano chiaramente quale ridondanza e inefficienza causerebbe un simile modo di esprimere la datazione (in entrambi i casi la maggior parte delle celle rimane inutilizzata ma deve comunque essere presente, dato che la tabella ha un numero costante di colonne). Inoltre, con date esprimenti un intervallo temporale (da A a B, o prima di B, o dopo di A) le celle dovrebbero essere ulteriormente moltiplicate (anzitutto raddoppiate, per prevedere l'espressione di un termine iniziale e di uno finale), ingigantendo in tal modo i difetti già evidenti di una simile definizione. Di fatto, tale è la varietà delle espressioni relative alla datazione che talora (per esempio nei *cd-rom* come il *Thesaurus Linguae Graecae* realizzati secondo le specifiche del *Packard Humanities Institute*) essa viene semplicemente espressa come testo (complicando in tal modo qualsiasi trattamento automatico, che dovrà interpretare il testo stesso per dedurre un valore numerico della data ai fini ad esempio di filtro o ordinamento cronologico), oppure ridotta alla mera indicazione del secolo. In molte altre iniziative editoriali di questo tipo in effetti si adottano simili prassi compromissorie, memorizzando ora solo l'indicazione testuale della data (con lo svantaggio di rendere assai difficile se non impossibile il loro trattamento automatico), ora solo un'approssimativa indicazione numerica (p.es. il numero del secolo, con lo svantaggio di perdere tuttavia le complesse sfumature cronologiche espresse dalla data nella sua forma testuale), ora entrambe (con i vantaggi ma anche gli svantaggi di entrambe le soluzioni, e con l'ulteriore penalizzazione di dover raddoppiare ogni volta l'espressione di una singola data, introducendo pericolose ridondanze). Nel progetto qui presentato invece si utilizza la potenza descrittiva di XML per esprimere la datazione in tutte le sue sfumature, senza nondimeno rinunciare a una sua rigorosa strutturazione capace di garantire la facilità di trattamenti automatici. A tale scopo tutte le varie espressioni di datazione come quelle esemplificate sopra sono 'scomposte' nei loro elementi atomici, ciascuno dotato di uno specifico ruolo semantico e definito da una precisa tipologia di valore: la datazione viene infatti espressa da un tipo complesso (*HistDate* nella figura), nel quale una data può essere definita in tre modi principali (tutti complementari):



- come punto preciso nel tempo: *pt* (*point*): ad esempio, “II secolo”.
- come intervallo temporale: *rng* (*range*): questo comprende a sua volta 1 o 2 elementi corrispondenti a due punti, *min* e/o *max*, che indicano i due limiti dell'intervallo. Ad esempio, *min* = “II secolo d.C.” indica una datazione a partire dal II secolo d.C.; *max* = “III secolo d.C.” indica una data non superiore al III secolo d.C.; *min* = “II secolo” e *max* = “III secolo” indica che l'opera è compresa fra II e III secolo.
- come ignota: *unk* (*unknown*). In questo caso, l'elemento *unk* può contenere una stringa qualsiasi che indica i motivi di questa datazione o altre notazioni.

Allo scopo di consentire trattamenti automatici, per tutti i valori temporali (cioè *pt*, *min* e *max*), il valore dell'elemento è sempre e solo un numero: eventuali attributi dell'elemento (dei meta-dati) indicano poi come il numero vada interpretato. Ad esempio, “II secolo a.C.” si scriverà come:

```
<pt cent="true">-2</pt>
```

dove *cent* è l'attributo *century* di valore booleano che quando vero indica che il numero va inteso come secolo, e il valore negativo indica l'era anteriore a Cristo. Altri attributi sono:

- *span*: indica che il valore va inteso estendersi all'anno successivo (p.es. “403/2 a.C.” per ovvie differenze nel calendario antico il cui inizio non coincide con quello gregoriano);
- *about*: vero se la data è indicativa (=‘circa’).
- *perhaps*: vero se la data è dubbia (=‘?’).
- *day*: indica il giorno del calendario gregoriano quando applicabile e presente.
- *month*: indica il mese del calendario gregoriano (sempre in cifre) quando applicabile e presente.

– *hint*: contiene opzionalmente una breve indicazione esplicativa relativa alla data in questione (p.es. “dopo la morte di Augusto” per spiegare una data come “posteriore al 14 d.C.”).

Come si vede, ogni elemento della datazione viene scomposto nei minimi termini, e a ognuno di questi termini viene riconosciuto un ruolo e un tipo specifico di dato (numerico, booleano, testuale, etc.). La struttura dell’elemento XML esprimente la datazione è suscettibile di ramificare in modi diversi (e complementari) a partire dalla radice (*date*), in modo da esprimere di volta in volta un singolo punto nel tempo (*pt*), un intervallo (*rng*) o una data del tutto ignota (*unk*); a sua volta, l’intervallo ramifica ulteriormente per contenere un minimo e/o un massimo (*min* e *max*), e ciascuno di questi sottoelementi (*pt*, *min*, *max*) contiene un valore numerico e tutti gli attributi eventualmente necessari alla sua interpretazione (se tale valore si riferisca a un secolo o a un anno, se sia solamente ipotetico o comunque dubbio, se si riferisca a una data a cavallo fra due anni del calendario gregoriano, etc.). In tal modo diviene dunque possibile esprimere ogni sorta di data secondo le convenzioni abituali, e nondimeno implementare dei semplici algoritmi per ricavare un numero che esprime univocamente la data e consente di effettuare un completo ordinamento o filtraggio cronologico; naturalmente, una data espressa secondo questo modello risulterà meno ovvia al redattore umano, ma appositi programmi schmeranno tale difficoltà facilitando l’immissione e la lettura delle date, e semplici trasformazioni (basate anche sul solo XSLT) saranno in grado di esprimere ogni datazione nella abituale forma testuale. Per citare a caso un esempio reale, si pensi a SEG 46,1531, la cui datazione è espressa come:

“367/366 B.C.? - ca. 150 A.D. (Roman copy)”

Una simile espressione viene così scomposta e strutturata in XML:

```
<date>
  <rng>
    <min span="true" perhaps="true">-367</min>
    <max about="true" hint="Roman copy">150</max>
  </rng>
</date>
```

dove:

- il primo termine dell’intervallo (“367/366 B.C.?”) viene espresso dall’elemento *min* dell’elemento *rng*, dotato degli attributi *span* (a indicare lo sfalsamento del calendario lungo due degli anni del calendario gregoriano: 367/6) e *perhaps* (a indicare una datazione dubbia, corrispondente al punto interrogativo) entrambi con valore booleano di *vero*, mentre il valore negativo -367 indica l’anno avanti Cristo.
- il secondo termine dell’intervallo (“ca. 150 A.D. (Roman copy)”) è indicato dall’elemento *max* dell’elemento *rng*, dotato dell’attributo *about* = *vero* che corrisponde al “ca.” dell’espressione originale, e del valore positivo 150, a indicare il 150 dopo Cristo. Inoltre l’attributo *hint* specifica l’indizio che ha indotto l’editore alla definizione di questo secondo termine cronologico (“Roman copy”).

Questi due elementi insieme definiscono con precisione l’intervallo cronologico citato, con tutte le sfumature dell’espressione testuale propria della scheda epigrafica del SEG, ma senza rinunciare a una rigorosa scomposizione dei suoi elementi e alla loro strutturazione, indispensabili per rendere agevole il trattamento automatico. Inoltre non si genera alcuna ridondanza, dato che sono presenti solo gli elementi necessari all’espressione della datazione esemplificata. Per salvaguardare comunque semplicità e apertura del modello, il redattore può comunque definire una simile datazione senza alcuna consapevolezza della sua struttura (nella fattispecie del codice XML) avvalendosi di un’apposita interfaccia grafica, illustrata qui sotto e relativa al programma redazionale per l’immissione dei dati:

Datation

Type: range

367/366 a.C. ? - c. 150 p.C. (Roman copy)

From (terminus a quo)

About (from)	False
Century (from)	False
Day (from)	0
Hint (from)	
Month (from)	0
Perhaps (from)	True
Span (from)	True
Value (from)	-367

To (terminus ad quem)

About (to)	True
Century (to)	False
Day (to)	0
Hint (to)	Roman copy
Month (to)	0
Perhaps (to)	False
Span (to)	False
Value (to)	150

Value (to)
Numeric value for year or century, negative if B.C.

Cancel OK

interfaccia grafica per l'immissione di una data

Come si vede, il redattore non deve che selezionare graficamente il tipo di datazione (casella `Type` in alto), in questo caso un intervallo cronologico (*range*), e riempire ove necessario le caselle relative ai due termini dell'intervallo stesso (`From` e `To`), mentre il programma si incarica di convertire questo *input* in codice XML formato secondo le specifiche descritte sopra. Nello stesso tempo si ha la possibilità di controllare la propria immissione ottenendo in tempo reale una descrizione testuale della data secondo il formato abituale delle edizioni tradizionali (cfr. la casella superiore riquadrata, dove appare l'espressione testuale completa relativa alla datazione appena specificata), generata automaticamente dal programma stesso. Si può peraltro osservare che la forma qui presentata al redattore è solo una delle infinite variazioni che si potrebbero adottare in rapporto ai propri intenti: ad esempio, sarebbe ugualmente possibile presentare la data in formati più compatti, esprimendo magari solo il secolo, o solo l'anno, con o senza le varie indicazioni accessorie, etc.: un unico contenuto può cioè incarnarsi in forme virtualmente infinite. In alternativa inoltre il redattore può semplicemente digitare la data come testo con tutte le sfumature del caso, lasciando al programma di *data entry* il compito di analizzare questo testo scomponendone gli elementi secondo lo schema appena illustrato.

In tal modo, il *software* realizzato *ad hoc* si frappone come intermediario fra il redattore (che può operare in modo molto più simile a quello tradizionale) e il codice XML automaticamente generato, che può quindi raggiungere la complessità desiderata (necessaria ai programmi di consultazione e trasformazione) senza ripercussioni negative sui processi redazionali. Un ulteriore vantaggio di tale approccio è offerto dalla completa opacità del codice XML rispetto al redattore, sicché diviene possibile alterare la struttura del primo (ad esempio a scopo di ampliamento o specializzazione) mantenendo inalterata l'interfaccia grafica preposta alla sua generazione: eventuali cambiamenti dello schema dell'informazione potranno quindi essere gestiti completamente dal programmatore senza alcun effetto visibile sugli strumenti operativi dei redattori, contribuendo all'apertura e all'espandibilità di un tale modello.

Naturalmente questo è solo un piccolo esempio, ma sufficiente a illustrare come l'informazione originaria venga scomposta fino al suo atomo, e ogni elemento risultante dalla scomposizione sia collocato

all'interno di una struttura rigorosa e fortemente ramificata, si da adattarsi alla complessità del materiale senza generare ridondanze e mantenendo una facile disposizione al trattamento automatico. Nello stesso tempo, redattore e utente finale sono schermati dalla complessità della banca dati XML, che giace sempre dietro le quinte e di cui entrambi non hanno alcuna consapevolezza: appositi programmi provvedono a trasformare l'*input* del redattore in codice XML, o all'inverso codice XML in forme dal contenuto e formato abituali per l'utente dell'edizione. Su scala ben più grande un tale paradigma viene implementato da una completa applicazione realizzata nell'ambito del progetto e capace di esporre un'interfaccia grafica amichevole e di facile utilizzo a redattori del tutto inconsapevoli del codice XML sottostante da esso generato o modificato; i redattori non devono che operare su una serie di schede elettroniche proprio come se dovessero riempire un modulo prestampato, del tutto indipendente dalla struttura e dal formato che i dati da loro immessi assumeranno effettivamente nella banca dati. Quanto emerge infine da questo esempio è la generale astrattezza e purezza semantica dei dati memorizzati, che sono del tutto indipendenti da una forma specifica e mirano esclusivamente alla loro completa strutturazione e definizione: non si tratta infatti qui di memorizzare una datazione secondo alcuna convenzione formale specifica (come appunto "367/366 B.C.? - ca. 150 A.D. (Roman copy)"): quanto l'operatore immette sono solo numeri e attributi deputati alla loro interpretazione, e ogni forma in cui essi sono presentati deriva da una trasformazione effettuata internamente dal programma. Ciò realizza una completa separazione tra forma e contenuto, sicché quest'ultimo risulta suscettibile di assumere qualsiasi veste l'editore ritenga opportuna per il suo pubblico, *medium* e formato di edizione. Allo stesso tempo, si è ben lungi da una semplice marcatura di un testo epigrafico, e assai più vicini a una vera e propria banca dati, definita solo dalla struttura più adatta ad esprimere ogni loro aspetto semantico. Con una semplice metafora, si potrebbe assimilare l'edizione cartacea meramente trasposta su *medium* elettronico a un blocco monolitico dove non è possibile isolare alcuno dei suoi componenti, e il testo marcato allo stesso blocco, però frammentato in modo da distinguere i suoi principali elementi costitutivi; rispetto a quest'ultimo caso, il modello qui illustrato opera un'ulteriore frammentazione degli elementi, giungendo idealmente sino al punto in cui essi non risultino ulteriormente (o convenientemente) frazionabili, e soprattutto scompone il blocco così frammentato rimuovendo i frammenti e raccogliendoli in contenitori diversi a seconda della loro natura. Sarà poi compito del *software* recuperare i vari elementi dai diversi contenitori per ricostruire il blocco originario, ma anche qualsiasi altra forma a partire dagli stessi mattoni di base.

Naturalmente questo comporta un notevole sforzo per la realizzazione di tutti gli adeguati strumenti *software*, ma un tale approccio, ispirato anzitutto dalle mie esperienze di analista e programmatore specie nel campo della ideazione e creazione di *corpora* testuali elettronici (una delle prospettive cui accennavo all'inizio dello scritto), può risultare particolarmente fruttuoso sul lungo termine, dato che una volta realizzati tali strumenti essi potranno essere alla base dell'edizione di qualsiasi altro *corpus*: il ricco apparato di generatori e trasformatori di codice XML potrà essere riusato *in toto*, e la separazione di redattori e utenti finali dalla banca dati garantirà in ogni fase redazionale una notevole libertà di variazione delle forme e ampliamento e specializzazione dei contenuti. Peraltro, la filosofia alla base di tale progetto consente intrinsecamente l'interoperabilità con progetti del tutto diversi, nella misura in cui i suoi contenuti, in sé privi di una forma specifica, sono per principio destinati a trasformazione tanto per una semplice pubblicazione quanto per una completa conversione in altri formati.

Fasi redazionali

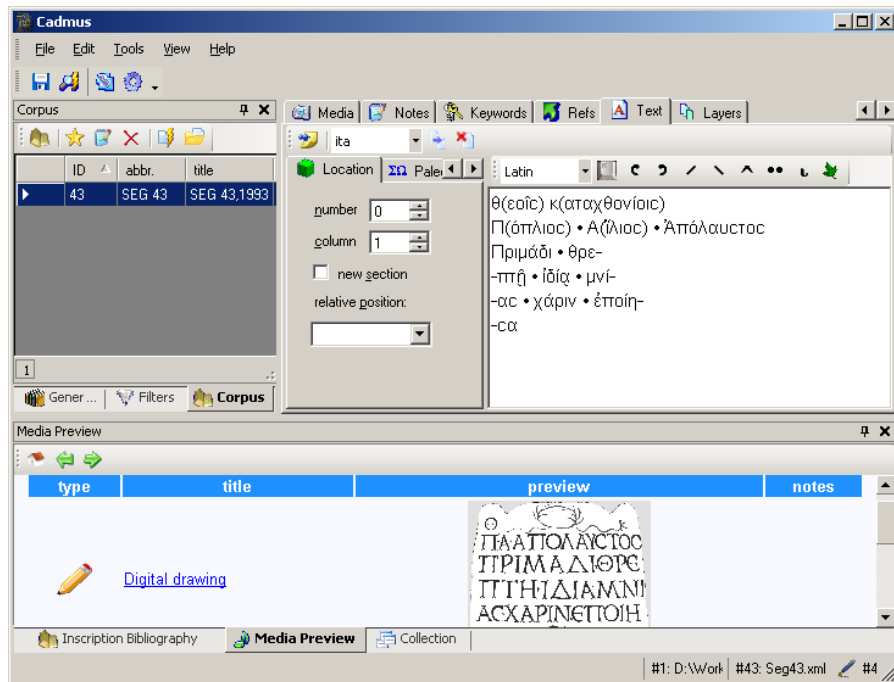
Come si è detto, il progetto qui proposto è il frutto di varie evoluzioni di un nucleo originale e prodotto di un insieme di applicativi *software* creati *ad hoc* per la sua concreta realizzazione; proprio la concretezza di un simile progetto costituisce la prima verifica dei suoi principi teorici: non si tratta qui infatti di presentare un'astratta proposta relativa a un possibile futuro, senza averla ancora messa a confronto con la realtà dei problemi connessi alla sua fattiva implementazione, ma di offrire le prime riflessioni e i primi risultati di un lavoro effettivamente in corso di realizzazione, a cominciare dai suoi strumenti *software*. In questo ambito può essere utile proseguire l'illustrazione di altri aspetti salienti del progetto con una breve analisi delle principali fasi redazionali in cui esso dovrà articolarsi, sì che anche i primi ne risultino meno astratti.

Immissione dei dati e codifica del testo

Il materiale più importante per l'edizione è poi naturalmente costituito dal testo dell'iscrizione, assieme a tutta la varia e ricca informazione connessa a una scheda epigrafica tradizionale. Anche in questo caso la filosofia generale del progetto viene rispettata adottando una complessa strutturazione dei dati, che su scala molto più larga opera secondo i principi già visti in piccolo per l'esempio relativo alla datazione. Non è questo il luogo per una dettagliata discussione relativa alla strutturazione della

banca dati epigrafica XML, che richiederebbe peraltro eccessivi tecnicismi; quanto è utile rilevare in termini generali si può comunque sintetizzare con due ulteriori osservazioni, relative alla codifica del testo e al rapporto fra il testo dell'iscrizione e quello dell'edizione.

Nel modello qui presentato, a monte del processo redazionale si trovano, oltre che ovviamente i documenti editi, i materiali multimediali da essi tratti (es. fotografie digitali) e quelli bibliografici ad essi relativi. Tutti i dati che si desidera inserire nell'edizione sono memorizzati al loro posto e secondo le loro specifiche esigenze all'interno di una articolata struttura XML contenuta in file di testo, cui si associano esternamente tutte le risorse multimediali. Per gestire l'inserimento e la scomposizione dei dati in queste strutture si utilizza un apposito *software* che scherma i redattori dalle complessità della codifica sottostante, consentendo loro di digitare testo come in un applicativo di videoscrittura o inserire ogni dato con l'ausilio di un'interfaccia grafica semplice e modulare. L'illustrazione anche sommaria di questo applicativo andrebbe oltre gli scopi di questo scritto, sicché mi limiterò qui a presentare un'immagine della sua interfaccia a scopo puramente esemplificativo¹³.



L'immagine mostra uno dei numerosi pannelli di immissione dei dati, relativo al testo epigrafico (in alto a destra), assieme a pannelli elencanti il contenuto del *corpus* (a sinistra), l'elenco delle risorse multimediali associate (in basso), etc. In generale, per l'utente si tratta di riempire i vari campi di una scheda epigrafica virtuale, articolata visivamente in una serie di pannelli ciascuno relativo a un dato tipo di informazioni (paleografiche, testuali, linguistiche, filologiche, storiche, archeologiche, multimediali, etc.); la particolare struttura del programma consente ad esso di crescere assieme all'edizione, dato che è facile aggiungere nuovi pannelli per consentire l'immissione di nuovi tipi di contenuti. Ogni modifica effettuata in un qualsiasi campo della scheda epigrafica presentata all'utente viene immediatamente codificata e inserita nella struttura XML del file in uso, senza che l'utente abbia alcuna specifica consapevolezza di come questo avvenga, dato che il programma si incarica di leggere i dati codificati per visualizzarli nella scheda, ed effettuare l'operazione inversa per ogni nuovo inserimento. In tal modo si evita di dover sottoporre i redattori a lunghe e complesse fasi di addestramento, necessarie magari in altri tipi di progetto dove l'utente stesso è deputato a inserire *tag* XML accanto al testo, e si sottopone inoltre ogni immissione a una immediata convalida che riduce la possibilità di errori. In questa fase l'opera consiste semplicemente di un qualsiasi numero di file XML, ciascuno dei quali contenente un gruppo di epigrafi con tutti i dati ad esse relativi; uno o più componenti del

¹³ Al momento il *software* per l'inserimento dei dati è realizzato con applicativi autonomi (scritti in linguaggio C#) destinati all'uso su singoli postazioni utente, ciascuno dei quali tipicamente lavora a uno dei numerosi *file* che compongono il *corpus*. In future evoluzioni del progetto si potranno comunque realizzare programmi ospitati su *server* dedicati per consentire l'immissione remota di dati via *web*.

personale della redazione inserisce questi dati (o almeno la parte minimale di essi¹⁴) utilizzando il programma e lavorando di volta in volta in un file specifico, senza necessariamente avere accesso all'insieme di tutti i file che costituiscono il *corpus*; una volta terminata la fase di immissione è dunque operazione banale raccogliere nuovamente insieme tutti i file all'interno del *corpus*.

Trattandosi di un campo dove le soluzioni messe in opera da altri progetti sono assai varie, è poi utile soffermarsi sul trattamento di testi basati su sistemi di scrittura diversi dall'alfabeto latino, come qui per il caso del greco. Malgrado la sempre maggiore diffusione di standard più recenti nati per codificare testi nelle varie lingue del mondo renda oggi meno problematico il loro trattamento, simili testi offrono infatti ancora difficoltà, e pongono comunque gli editori dinanzi a delle scelte. È ormai evidente a chiunque consulti pagine *Internet* o utilizzi sistemi di sviluppo *software* nel mondo dell'informatica personale che lo standard di fatto per la codifica di testi in lingue non basate su alfabeti latini è costituito oggi da *Unicode*, che a partire dalla sua versione 2.0 si è peraltro allineato all'altro grande standard concorrente dopo la caotica epoca delle innumerevoli codifiche a 8 bit, ISO 10646. Quale che sia la posizione teorica assumibile dinanzi allo standard, che come tutti gli standard di successo deve la sua diffusione anche a compromessi pratici con la situazione di fatto, quanto non si può ignorare è appunto la sua ormai universale diffusione e adozione in tutte le piattaforme più recenti, dai *personal computer* sino ai dispositivi portatili, e per conseguenza in tutti i principali sistemi di sviluppo di *software* a essi destinato. Per garantire il facile scambio dei dati e la longevità che deriva dall'adozione degli standard più diffusi oltre che per motivi pratici connessi alla creazione dei programmi redazionali oggi appare dunque quasi obbligata la scelta di *Unicode* per la codifica di testo greco o comunque non basato sull'alfabeto latino. Nondimeno rimangono alcune oggettive difficoltà pratiche, lamentate da diversi editori operanti in questo settore, che si possono riassumere in due punti principali: l'effettiva mancanza di caratteri di uso specialistico in campo epigrafico e filologico da un lato, e la non rara difficoltà di digitazione di testo *Unicode* utilizzando sistemi operativi o applicativi più o meno recenti.

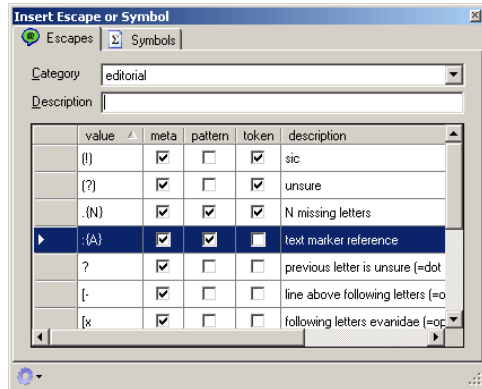
Rispetto al primo punto, benché la situazione stia lentamente migliorando con le successive versioni dello standard¹⁵, è innegabile che l'attuale copertura non risulti comunque sufficiente per tutte le esigenze di un'edizione filologica tradizionale; inoltre, fatto talora persino più rilevante sul piano pratico, risulta ancora difficile reperire *font* contenenti tutti i caratteri utili a questo scopo, benché magari già presenti nello standard *Unicode*. Rispetto al secondo punto poi la grande varietà di piattaforme, applicativi di videoscrittura e *font* utilizzati per i testi greci rende spesso difficile individuare una soluzione capace di soddisfare almeno la maggior parte dei potenziali redattori, ciascuno dei quali tende a mantenere le proprie abitudini di digitazione. Non è anzi raro trovare chi ancora oggi continui a digitare testi greci avvalendosi di *font* dalla codifica del tutto arbitraria, dando quindi luogo a inevitabili problemi di portabilità ogni volta che il testo debba essere trasferito da una macchina all'altra. Una simile situazione talora caotica non può però giustificare la semplice prosecuzione di una simile prassi, che nuoce ovviamente ancor più in progetti editoriali di ampio respiro. Per il progetto qui presentato dunque si è scelta la codifica *Unicode*, ma con l'adozione di alcune strategie utili a superare entrambe le difficoltà citate.

In primo luogo, rispetto alla mancanza di caratteri non sarebbe ovviamente pensabile di definire arbitrariamente dei codici di carattere da riservare a questo scopo, prassi che contravverrebbe alla stessa natura di standard della codifica che si è scelto di adottare proprio in virtù della sua diffusione. D'altra parte, anche in relazione alla rarità di *font* contenenti tutti i caratteri necessari, o alla semplice difficoltà di digitare determinati caratteri di uso relativamente meno frequente (si pensi ad es. a tutto l'apparato di simboli filologici o epigrafici pure previsti da *Unicode* ma difficilmente presenti in *font* o inseribili da tastiera), si è adottata qui una semplice e assai diffusa procedura, basata su una sequenza di *escape*: è infatti sufficiente riservare un carattere dello standard *Unicode* facilmente accessibile da tastiera ma difficilmente presente in un testo di questo tipo a un uso speciale, per introdurre cioè la definizione di caratteri assenti dal *font* e/o dallo standard o semplicemente di difficile inserimento, per superare entrambe le difficoltà e offrire inoltre la possibilità di distinguere annotazioni metatestuali dal testo vero e proprio qualora sia opportuno inserire queste ultime assieme al testo stesso. Ad esempio, riservando alla notazione di tali sequenze speciali i caratteri @ e \$ per delimitarne i confini, nel testo si possono facilmente inserire caratteri presenti nello standard ma magari assenti dal *font* o comunque

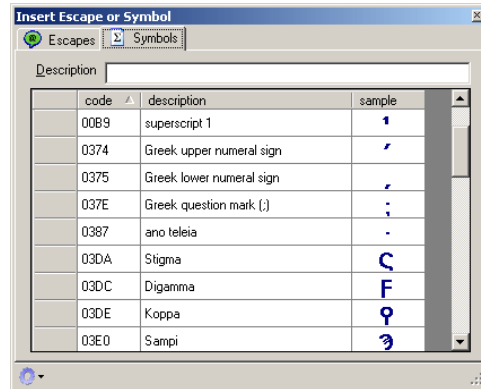
¹⁴ Benché la scheda epigrafica nella sua forma attuale comprenda centinaia di campi per soddisfare tutte le esigenze per cui il progetto è stato sviluppato, la maggior parte di essi rimane opzionale sicché l'utente è sempre libero di inserire solo i dati essenziali, rimandando magari ad altro momento il loro completamento.

¹⁵ A questo proposito basterà ricordare che con l'estensione dell'*encoding space* a piani multipli nel passaggio da 16 a 21 bit *Unicode* ha riservato un intero piano (*Supplementary Multilingual Plane*) proprio a ospitare tutti i caratteri relativi a sistemi di scrittura arcaici o obsoleti, di interesse per gli studiosi del settore.

difficilmente inseribili da tastiera, o anche parzialmente o del tutto assenti dallo standard: così si potrebbe rappresentare un *hedera distinguens* con @hedera\$, una palmetta con @palma\$, un numerale del sistema acrofonico attico come Π (=500) con @a500\$, un *chrismon* con @+2627\$ (dove + introduce il codice di carattere *Unicode*, magari però assente dal *font* che si è deciso di utilizzare) etc. Inoltre ciò rende possibile inserire notazioni metatestuali come *vacat* che la prassi come motivi di convenienza inducono a inserire con il testo stesso, e che dal testo stesso risulteranno comunque facilmente distinguibili proprio per la loro rappresentazione in simili sequenze (p. es. @vac.12\$). Il programma di immissione di dati è in grado di riconoscere e convalidare la sintassi di tutte le sequenze che l'utente è libero di definire in un apposito repertorio, assieme a ogni dettaglio della loro resa tipografica finale per il fruitore dell'edizione. Inoltre la sua interfaccia consente all'utente di inserire visivamente simboli anche presenti nello standard come nel *font* prescelto ma difficilmente digitabili da tastiera (cfr. le figure riportate qui sotto).



inserimento di sequenze di escape metatestuali



inserimento di simboli

Una volta memorizzate nel testo, tali sequenze non appariranno ovviamente come tali al fruitore dell'edizione, ma saranno automaticamente convertite dal programma di pubblicazione in qualsiasi altra forma, ad esempio attingendo simboli da altri *font* (anche non *Unicode*, il che è inevitabile per caratteri assenti dallo standard) esplicitamente rivolti alla pubblicazione, o modificando il formato tipografico come richiesto dall'editore per ogni singola sequenza. In tal modo diviene possibile rappresentare qualsiasi carattere utile all'edizione sia esso presente o meno nello standard e/o nel *font* prescelto (che non deve necessariamente corrispondere ad alcun *font* scelto per la pubblicazione finale), senza introdurre alcun codice di carattere arbitrario.

Quanto all'immissione di testo, una volta superati i problemi relativi alla disponibilità dei caratteri rappresentabili con il *font* prescelto si tratta solo di fornire adeguato supporto *software* per la digitazione, rispetto alla quale sono disponibili due strategie complementari: utilizzando il programma di immissione dei dati si dispone direttamente di un'interfaccia creata appositamente per digitare in modo semplice testo greco e latino con tutti i relativi diacritici, con l'ausilio (a scelta) di tastiera, *mouse* e tastiere virtuali visualizzate direttamente sullo schermo. In alternativa¹⁶, gli stessi strumenti *software* realizzati per il programma di immissione sono disponibili direttamente dall'interno di applicativi di videoscrittura come *MS Word*, per consentire di digitare testo *Unicode* al suo interno o di convertire automaticamente in *Unicode* testo digitato utilizzando il proprio *font* preferito, quale che sia la sua codifica.

Naturalmente, in virtù della separazione tra forma e contenuto, l'adozione di una determinata codifica testuale per il greco non implica certo che essa debba corrispondere né a quella di una determinata fase redazionale (ad es. l'immissione dei testi da parte dei redattori), né tantomeno all'effettiva pubblicazione dei dati per l'utente finale. In entrambi i casi, gli stessi strumenti *software* di conversione di formato realizzati *ad hoc* consentono di convertire il testo automaticamente quale che sia la sua codifica originaria, arricchendolo eventualmente di tutta la formattazione tipografica relativa al formato di *output* prescelto (ad es. RTF, HTML+CSS, PDF, etc.). In tal modo si ottiene una completa indipendenza dalla forma che tradotta in termini redazionali implica anche una indipendenza da specifiche risorse tipografiche, prima di tutto ad

¹⁶ Questa soluzione si propone in particolare per il *web*, nella prospettiva di un trasferimento del programma dal *computer* dell'utente a un *server* cui l'utente si collega per immettere dati remotamente. In tal caso l'utente può aver già digitato il testo nel suo applicativo di videoscrittura preferito, con il suo *font* preferito, sia esso *Unicode* o meno, ed è in grado di copiarlo direttamente nella scheda epigrafica che provvederà alla sua riconversione in *Unicode*.

esempio da specifici *font* di caratteri, dato che è sempre possibile generare dinamicamente a partire da *Unicode* testo pronto per essere visualizzato con uno o più tipi di carattere scelti in modo del tutto arbitrario. Un apposito modulo *software*, sviluppato con finalità ancora più ampie di questo progetto, è deputato alla conversione di testo attingendo a un qualsiasi numero di *font* di caratteri, *Unicode* o meno, dotati o meno di caratteri sovrapponibili. In breve, tale modulo si basa su un modello di astrazione assai maggiore rispetto alla semplice equivalenza di sequenze di codici di carattere; per meglio comprendere la sua natura si immagini il seguente scenario:

- un testo di origine è codificato secondo il *font* A, e occorre generare il medesimo testo in un *font* B la cui codifica è diversa da A (ad es., il codice 187 corrisponde a *alpha* + acuto + iota sottoscritto in A ma a omega + iota sottoscritto in B¹⁷, sicché non è ovviamente possibile una mera sostituzione di B ad A).
- il testo di origine comprende una *alpha* con macron e accento acuto, realizzata con 3 caratteri diversi, uno per *alpha* seguito da due caratteri sovrapposti al precedente per i due diacritici.
- il *font* B non usa caratteri sovrapponibili (come A) ma solo caratteri dove i diacritici sono precombinati con le lettere in modo da esaurire tutte le combinazioni. Tuttavia, esso non dispone di un singolo carattere per la combinazione *alpha* + acuto + macron del testo originale: tutto ciò che ha è solo *alpha* oppure *alpha* + acuto.

Innanzitutto non sarebbe possibile utilizzare un semplice meccanismo di conversione basato sulla mera corrispondenza di sequenze di caratteri (una sequenza di caratteri del *font* A come *alpha* + acuto + *macron* corrisponde a un'altra sequenza nel *font* B, come *alpha* & acuto¹⁸), dato che ciò implicherebbe la necessità di prevedere e definire a priori tutte le possibili combinazioni che costituiscono una sequenza nel *font* A, che nel caso dell'esempio potrebbero essere:

alpha + acuto + *macron*
alpha + *macron* + acuto

Se si pensa al numero di diacritici che un carattere greco può contenere (ad es. *alpha* + spirito dolce + accento acuto + iota sottoscritto etc.) è facile comprendere come si finirebbe subito per dover prevedere un numero eccessivo di combinazioni: ad esempio, con soli 3 diacritici si potrebbero avere 27 combinazioni per ogni lettera.

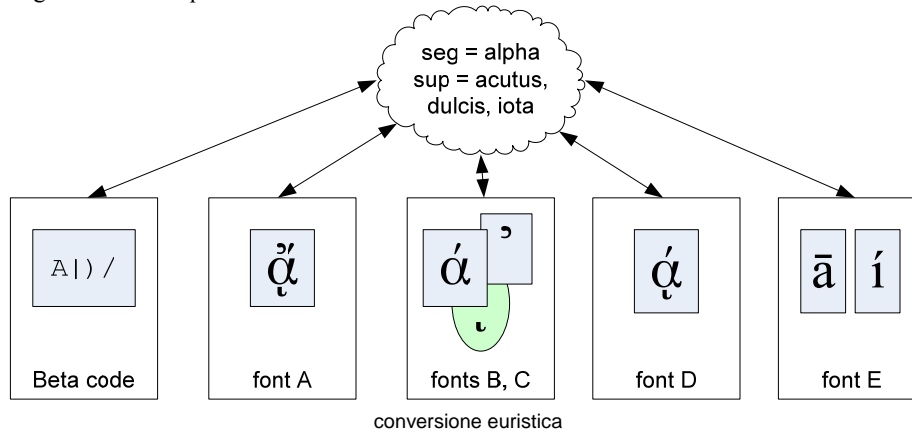
Inoltre, i due *font* A e B non si corrispondono perfettamente, dato che A ha la capacità di esprimere il diacritico *macron* che invece manca nel *font* B; in ogni caso, dovendo rappresentare il testo originario nel *font* B sarà preferibile una conversione anche parziale piuttosto che nessuna conversione: in altri termini, convertire *alpha* + acuto + *macron* in *alpha* + acuto sarà preferibile a non convertire affatto l'intero carattere dell'originale semplicemente perché un diacritico su tre manca di corrispondenza nel *font* B.

Allo scopo di risolvere tutti questi problemi e ottenere di volta in volta sempre il risultato migliore viene adottato un meccanismo di conversione euristico, che non opera con equazioni predefinite di sequenze di caratteri (tali caratteri di A corrispondono a tali caratteri di B) ma piuttosto con delle entità astratte, per cui un carattere come α è in realtà analizzato come un insieme di quattro diverse entità: *alpha*, spirito dolce, accento acuto, iota sottoscritto. Ogni volta che il convertitore trova un carattere nel *font* A lo analizza e scompone in entità avvalendosi di una apposita tabella XML che descrive ogni carattere di A nelle relative entità; quindi il convertitore esamina la tabella descrittiva di B alla ricerca di tutti i suoi caratteri (o combinazioni di più caratteri) capaci di rappresentare il maggior numero di entità trovate in A: se tutte le entità vengono trovate si otterrà una conversione perfetta, altrimenti si otterrà comunque il migliore risultato possibile in rapporto alle possibilità del *font* B. In tal modo un unico modulo *software* consente di gestire la conversione tra *font* con diverse codifiche, analizzando e scomponendo le loro sequenze di caratteri in gruppi di entità e quindi all'inverso ricomponendo delle sequenze di caratteri in uno o più *font* di destinazione a partire dalle entità stesse. Un tale convertitore è quindi euristico nella misura in cui non opera secondo equivalenze predefinite, ma di volta in volta trova la soluzione più adatta a rappresentare la conversione migliore da A a B.

¹⁷ Ovviamente questo può accadere solo quando i due *font* non siano *Unicode*, come di fatto avviene molto spesso nel trattare con *font* di greco antico: chiunque potrà aver fatto esperienza della difficoltà di scambiare testi greci antichi in formato elettronico (p.es. documenti di *Word*) senza disporre del medesimo *font*; si tratta di un altro degli aspetti per cui la diffusione di *Unicode* specie in Italia risulta ancora limitata.

¹⁸ In questo esempio uso il carattere + a indicare due caratteri distinti disposti in sequenza, e il carattere & a indicare un unico carattere che contiene due o più entità (ad esempio un carattere che combina *alpha* con accento acuto è notato come *alpha* & acuto).

In realtà la conversione effettuata da questo modulo *software* è assai più complessa, dato che non opera solo da *font* a *font* ma anche da *Beta code* (il formato a base ASCII caratteristico di storici *corpora* digitali codificati secondo standard definiti dal *Packard Humanities Institute* come il celebre *Thesaurus Linguae Graecae* dell'Università di Irvine) verso altri formati (ad es. (X)HTML o RTF); tuttavia il nucleo euristico rimane il medesimo per ogni tipo di conversione. Peraltro, la natura 'astratta' del convertitore (ovvero non diretta da formato a formato, ma mediata dalla scomposizione in entità) consente di convertire testo in modi anche più vari, per esempio allo scopo di traslitterare il greco in alfabeto latino: sarà infatti sufficiente descrivere l'alfabeto latino di destinazione negli stessi termini di entità del formato sorgente per ottenere una completa traslitterazione. A scopo esemplificativo, la figura seguente illustra qualche casistica astratta:



Rispetto a un astratto gruppo di entità (alcune delle quali 'segmentali' e altre 'sovrasedimentali', cioè sovrapposte a un'entità segmentale precedente) come *alpha*, accento acuto, spirito dolce e iota sottoscritto, le diverse rappresentazioni possono essere:

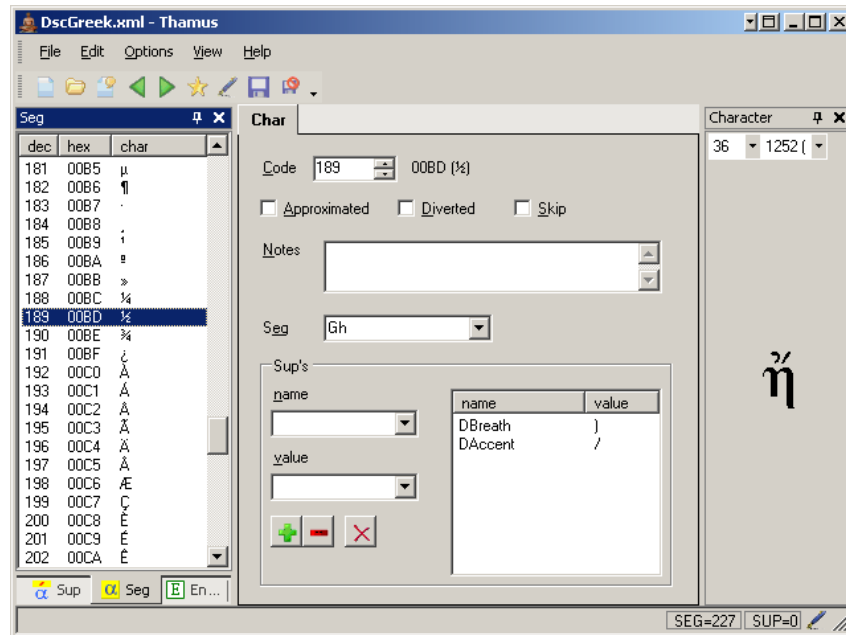
- in *Beta code* A |) / (dove A = *alpha* e gli altri caratteri corrispondono ai diacritici);
- in un ipotetico *font* A dotato di un unico carattere precombinato, il carattere con tutti i diacritici;
- in due ipotetici *font* B e C, i caratteri per *alpha* e spirito dolce sovrapponibile tratti dal *font* B e il carattere per iota sottoscritto tratto dal *font* C (evidentemente perché assente dal *font* B): come si vede, l'*output* del convertitore può riassembleare una sequenza di caratteri attingendo entità da più *font* diversi;
- in un ipotetico *font* D mancante del carattere (o dei caratteri) per *alpha* con accento acuto, spirito dolce e iota sottoscritto, il carattere che più si approssima a quello desiderato (mancante di spirito dolce);
- in un ipotetico *font* E, latino e non greco, il risultato è una vera e propria traslitterazione.

Come si vede, il convertitore ottiene caso per caso il risultato migliore in rapporto alle risorse tipografiche disponibili, che peraltro possono essere accresciute in ogni momento semplicemente aggiungendo ai *font* di *output* altri contenenti i caratteri necessari. Inoltre, l'astrazione relativa allo stadio intermedio di entità frapposto tra il formato sorgente e quello di destinazione offre numerosi vantaggi pratici: anzitutto, descrivere un singolo *font* in entità (ad es. specificando che al codice di carattere 165 corrisponde *alpha* + spirito dolce + accento acuto) è un sistema molto più vicino al modo di pensare del redattore umano; inoltre una singola descrizione di un *font* può essere riutilizzata senza alcuna modifica in qualsiasi tipo di conversione, nella misura in cui essa non si riferisce a un qualsiasi tipo specifico ma si limita a offrire la descrizione completa di un *font* (una volta descritto il *font* A, la sua tabella di descrizione può essere usata indifferentemente per tutte le conversioni che coinvolgano A sia come *input* che come *output*: da A a B+C, da C ad A, da B ad A+C, etc.). Più descrizioni possono poi essere combinate in modo da consentire l'uso di *font* multipli in *output* rispetto a un unico *font* in *input*; infine, è facile per il convertitore ignorare i dettagli implementativi del singolo *font* riguardo alla natura sovrapponibile o meno dei suoi caratteri, dato che una volta operata la scomposizione in entità tale distinzione perde ogni rilevanza.

In conclusione quindi il convertitore qui realizzato consente di trasformare testo codificato in *Unicode* (sia esso latino o greco) in qualsiasi altro formato attingendo a un qualsiasi numero di *font* scelti arbitrariamente, siano essi compatibili o meno con *Unicode*, e dotati o meno di caratteri sovrapponibili, adattandosi di volta in volta alle risorse a disposizione per ottenere sempre il miglior risultato. Questo consente di non vincolare la banca dati all'uso di una qualsiasi specifica risorsa

tipografica (un *font* piuttosto che un altro), e di adattare di volta in volta l'*output* al formato e al *medium* desiderati, ancora una volta in conformità ai principi generali del progetto editoriale e alla sua profonda vocazione alla trasformazione.

Anche in tal caso peraltro la complessità relativa alle risorse XML necessarie alle conversioni testuali (le tabelle descrittive dei *font*) viene schermata per il redattore da *software* realizzato allo scopo, che offre la possibilità di creare intere tabelle tramite un'interfaccia grafica che genera automaticamente il codice XML relativo all'*input* dell'utente. Un'idea di tale processo può essere offerta dalla figura riportata qui sotto, riguardante l'interfaccia principale del programma per l'*editing* delle tabelle XML di descrizione di *font*.



programma redazionale per l'*editing* di tabelle XML di descrizione di *font*: nell'esempio si mostra la descrizione del carattere 189 del diffuso *font Greek* di P.Gentry e A.Fountain, articolata nella sua parte 'segmentale' (*seg*: lettera eta) e 'sovrasegmentale' (*sup*'s: spirito dolce e accento acuto); un'anteprima del carattere descritto appare sul pannello di destra, mentre da quello di sinistra è possibile navigare all'interno dell'intero insieme di caratteri.

Stratificazione del testo e ortografia

Si è dunque visto quale sia la codifica adottata per il testo greco (o latino) della banca dati epigrafica XML, e come in ogni caso l'uso di *Unicode* (con il comodo espediente degli *escape* per superare alcune difficoltà pratiche nella sua applicazione) sia poco più di un dettaglio relativo alla sua implementazione interna, dato che appositi strumenti *software* consentono da un lato l'immissione di testo greco o latino nella forma familiare al redattore per poi generare automaticamente *Unicode*, e dall'altro la trasformazione di *Unicode* in un qualsiasi altro formato di *output*. In questo senso, in linea teorica il redattore come l'utente finale non hanno alcuna consapevolezza del contenuto nella sua struttura interna, dato che ogni sua visualizzazione implica intrinsecamente una sua trasformazione.

Il secondo esempio qui scelto per illustrare in modo meno astratto l'effettiva realizzazione della rigorosa separazione tra l'unico contenuto e le sue molteplici forme è costituito non più dalla codifica del testo, ma dalla sua struttura all'interno della banca dati XML. Tutti gli esempi sinora illustrati hanno mostrato come sia costante la cura di mantenere la neutralità del contenuto e la sua sistematica scomposizione e strutturazione nei suoi elementi atomici, si da offrire poi possibilità di una loro ricomposizione in modi e forme virtualmente infiniti. Anche la strutturazione del testo quindi non fa eccezione, e si riferisce in particolar modo alla soluzione dei problemi relativi all'ortografia del testo epigrafico citati sopra, relativi al rapporto tra la grafia storica e quella convenzionale. Peraltro, a proposito dei limiti relativi alla marcatura del testo si è già visto come sia possibile associare a porzioni del testo svariatissime e virtualmente infinite annotazioni di ogni sorta di carattere (epigrafico, paleografico, archeologico, storico, prosopografico, filologico, linguistico, etc.); in questo senso, tenendo conto della necessità di mantenere una struttura facilmente trasformabile e sempre aperta all'espansione senza pregiudizio per il materiale già esistente, e ancor più della intrinseca limitazione di qualsiasi tecnologia di marcatura come XML (che non ammette sovrapposizioni), non sarebbe possibile inserire

tutti questi dati all'interno del testo originale sottoforma di marcatura. In conformità peraltro al principio della memorizzazione separata di ogni singolo dato nel suo specifico contenitore all'interno della grande struttura gerarchica ad albero alla base della banca dati, la soluzione qui adottata per consentire di associare senza limitazioni qualsiasi tipo di informazione a qualsiasi porzione di testo si può assimilare all'immagine di un libro, dove un qualsiasi numero di pagine è collegato all'unica costola che le tiene insieme. Fuor di metafora, la costola del libro è rappresentata dal testo originale così come appare sulla pietra, e le pagine sono costituite da qualsiasi tipo di dato metatestuale si voglia associare a una qualsiasi porzione di questo testo. Ogni pagina contiene un certo tipo di dato metatestuale, strutturato e codificato secondo i propri principi, e collegato per riferimento al testo originale. In tal modo, mentre il testo originale rimane inalterato e riflette in tutto l'ortografia 'storica' del documento, rimane possibile aggiungere in qualsiasi momento e senza alcuna limitazione nuovi 'strati' sopra questo testo, ciascuno contenente informazione specifica a un dato campo di interesse. Ciò vale anzitutto per il primo strato metatestuale, che definisco 'ortografico': nel testo di base (la costola del metaforico libro) si riportano esattamente le varie righe così come appaiono sulla pietra, con tutti gli eventuali 'errori' ortografici e la loro impaginazione originale, e la sola aggiunta di diacritici, integrazioni e scioglimenti di abbreviazioni¹⁹; qualsiasi 'correzione' tesa a ripristinare la nostra convenzionale ortografia standard rispetto a quanto appare sulla pietra viene invece ospitata a parte su un apposito strato ortografico. In tal modo si memorizza da un lato il testo del documento nella sua grafia storica, e dall'altro tutte le varianti normalizzate secondo la prassi ortografica comune; questo consente dunque di superare qualsiasi compromesso nella pubblicazione del testo, così come qualsiasi difficoltà pratica nell'uso dell'edizione: infatti il testo che si memorizza risulta in qualche modo 'doppio', sia nella sua forma originale che in quella normalizzata, pur evitando ridondanze nella misura in cui nello strato ortografico, complementare al testo di base, si annotano solo le forme che richiedono normalizzazione senza ripetere l'intero testo. All'atto della pubblicazione poi lo stesso utente dell'edizione può decidere di visualizzare il testo nella sua forma originale o normalizzata, o in una qualsiasi combinazione delle due (p.es. forma normalizzata nel testo e forma originale nell'apparato); inoltre, qualora effettui una ricerca all'interno del *corpus* sarà indifferente se ad essere cercata sia la forma nella sua grafia storica (ad es. ἐτύθησεν) o normalizzata (ad es. ἐποίησεν): in entrambi i casi infatti il programma di consultazione sarà in grado di rintracciare il testo in cui essa occorre, dato che può estendere la ricerca sia al testo di base (dove appare ἐτύθησεν) sia allo strato ortografico (dove appare ἐποίησεν). Infine un simile approccio apre la possibilità di ricerche avanzate combinando le informazioni dello strato ortografico con quelle del testo di base: ad esempio l'utente potrebbe cercare tutte le forme di ἐποίησεν con una data grafia itacistica, o tutte le parole che in qualsiasi testo presentino *v* per *oi*, etc.: è facile in questo senso comprendere le potenzialità di una simile edizione in campo paleografico e linguistico, tanto più che lo stesso programma di pubblicazione potrà generare in modo del tutto automatico indici estremamente dettagliati e completi di tutte le corrispondenze ortografiche, capaci di essere usati come di repertori di errori paleografici o fatti linguistici notevoli. Basterà infatti che il programma di pubblicazione ripercorra il testo di ogni epigrafe assieme al suo strato ortografico, raccogliendo e classificando le corrispondenze tra le due grafie, storica e convenzionale, per generare automaticamente elenchi completi e dettagliati destinati alla consultazione da parte dell'utente dell'edizione, sia essa digitale o cartacea.

Per consentire questi e altri usi avanzati lo strato epigrafico adotta opzionalmente una notazione 'operativa' dell'ortografia, non solo riportando la forma ortograficamente 'corretta', ma correndandola di tutte le operazioni necessarie per trasformare la forma storica in quella convenzionale. Appositi elementi (detti *msp* = *misspelling*) forniscono infatti per ogni parola la descrizione puntuale di ogni

¹⁹ Lo scopo di tale norma è di riportare un testo contenente solo le lettere presenti sulla pietra così come esse vi appaiono, senza alcuna correzione, dato che nel testo l'intervento editoriale si limita alle sole aggiunte (scioglimento di abbreviazioni e integrazioni), che possono essere facilmente eliminate con procedure automatiche. Tutte le altre notazioni (quelle cioè che tradizionalmente si inseriscono fra parentesi diverse da tonde o quadre) sono riportate in strati distinti (per la nozione di strato cfr. sotto). Pertanto, una forma integra come ἐτώ per ἐτών sarà scritta ἐτώ nel testo, e altrove sarà riportata la forma corretta ἐτών; invece, una forma mutila ἐτῶ[v] sarà notata così nello stesso testo, trattandosi non di correzione ma di integrazione. Questo tipo di prassi potrebbe apparire incoerente in casi come quello citato, dove la correzione si limita all'aggiunta (e non alla modifica) di una lettera; vi sono tuttavia buoni motivi per proporre questa convenzione di notazione distinta: anzitutto, non tutte le correzioni si limitano all'aggiunta di una o più lettere, sicché sarebbe incoerente distribuirne alcune nel testo e altre in uno strato distinto; in secondo luogo poi, la notazione separata di ogni correzione al testo consente di poterle indicizzare ciascuna in modo distinto, sì da poter effettuare ricerche al loro interno o creare indici di interesse grammaticale o paleografico. Sul piano concettuale infatti interventi editoriali come scioglimento di abbreviazioni (=testo aggiunto fra parentesi tonde) e integrazioni (=testo aggiunto fra parentesi quadre) sono mere notazioni alternative (*M.* = *Marcus*) o interventi di restauro di un testo che per accidenti materiali non è pervenuto intero (per quanto simili interventi possano essere problematici, ogni altra informazione ad essi relativa sarà comunque pertinenza dell'apparato); le correzioni editoriali invece si basano sull'osservazione di aspetti linguistici errati rispetto all'ortografia assunta come normativa dalla moderna filologia: il lapicida che nota con *I* quanto oggi si scrive con *EI* non fa altro che esprimere in forma scritta un suono della propria lingua corrente.

singola deviazione dall'ortografia standard presente nella forma originale. Ancora una volta è qui operante il principio della scomposizione e dell'isolamento di ogni elemento semanticamente distinto dell'informazione: non si riporta direttamente nel testo di base la forma corretta, il che condurrebbe alla commistione di due piani distinti di informazione (da un lato quanto è sulla pietra, dall'altro quanto corrisponde ai nostri standard ortografici e tutto quanto costituisce un intervento editoriale in senso più lato): piuttosto, si memorizza separatamente da un lato la forma originale (p.es. ἡρώνη), dall'altro (nello strato ortografico) la sistematica descrizione delle sue deviazioni dalla norma ortografica (η per εἰ nella prima sillaba, mancata notazione di ι nell'uscita del dativo). In altri termini, un elemento *msp* definisce le trasformazioni cui deve essere sottoposta la parola originale per ottenere la parola ortograficamente corretta. Dal punto di vista meccanicamente formale, una sequenza di caratteri può essere modificata per sostituzione, inserzione o rimozione di uno o più caratteri con altri; dunque sarebbe sufficiente prevedere queste tre operazioni di base per definire (tramite la loro combinazione) ogni possibile trasformazione. Si potrebbe allora pensare a un meccanismo per cui a una data sequenza di caratteri in qualsiasi posizione nella parola ne viene fatta corrispondere un'altra o zero. Definendo una sequenza con un numero indicante il carattere (1=primo, 2=secondo, etc.: simbolizzato con @) e un secondo numero indicante la sua estensione in caratteri (simbolizzato con *) si potrebbero quindi definire le seguenti operazioni di base²⁰:

- **sostituzione**: ad es. *b* con *v* in *inbide* > *invide*: operazione @3*1="v", ovvero 'sostituisci la sequenza definita dal carattere numero 3 con l'estensione di 1 carattere [cioè "b"] con la sequenza "v".
- **cancellazione**: ad es. *connstantia* > *constantia*: operazione @4*1="" (che sostituisce il secondo *n* con zero).
- **inserzione**: ad es. *cosul* > *consul*: operazione @3*0="n" (che inserisce *n* prima di *s* senza cancellare *s*).

Come si può constatare, un simile modello già prevede combinazioni delle operazioni di base (ad es. per *pena* > *poena* l'operazione @2*1="oe" condensa una sostituzione e una inserzione) in virtù del fatto di operare con sequenze di estensione variabile.

Tale modello risulta però imperfetto nella misura in cui lo scopo di queste trasformazioni non è solo quello di ottenere un'astratta sequenza di caratteri a partire da un'altra, ma anche di riflettere quanto più possibile nelle loro operazioni i fenomeni linguistici che ne sono alla base (in omaggio peraltro alla seconda delle due prospettive cui accennavo all'inizio di questo scritto, quella appunto dello studioso della lingua). In tal modo sarà possibile non solo generare automaticamente testo ortograficamente corretto a partire da quello presente sulla pietra, ma anche (e nello stesso tempo e a partire dagli stessi dati) generare automaticamente degli indici essenziali degli errori di un testo e della loro tipologia linguistica, utile a scopi paleografici e linguistici. Per fare un esempio banale, si pensi al caso citato sopra di *inbide* / *invide*, dove l'operazione di trasformazione è formulata in modo da evidenziare l'equivalenza di V e B secondo la ben nota tendenza del latino volgare alla convergenza dei due suoni [w] e [b] in [v]. Questo è anche il motivo per cui laddove sia teoricamente possibile più di una formulazione viene sempre scelta quella più aderente al modello linguistico: ad esempio, nel caso di *pena* si potrebbe ottenere *poena* in due modi:

- a) @2*0="e" (inserzione di *o* prima di *e*)
- b) @2*1="oe" (sostituzione di *o* con *oe*)

Sebbene a prima vista la soluzione (a) possa apparire più economica, è la soluzione (b) ad essere preferibile in virtù del fatto che formalmente esprime con più evidenza il fatto linguistico alla base della deviazione dalla norma ortografica, cioè la monottongazione di *oe* in *e*: nel caso di (a) infatti il rapporto è fra zero ed *e*, nel caso di (b) fra *e* ed *oe*.

Si pensi ora a una parola come *leriquus* per *reliquus*: astrattamente si potrebbe trasformare la prima forma nella seconda con le operazioni:

@1*1="r"; @3*1="l"

cioè la sostituzione di *r* a *l* e viceversa: da una simile formulazione risulterebbe già meno evidente la natura del fenomeno linguistico sotteso da questa forma, cioè una metatesi per cui non si tratta del cambiamento di due fonemi isolati (*r* in *l* da un lato e *l* in *r* dall'altro), ma dello scambio di posizione di due liquide in sillabe adiacenti. Se un simile inconveniente potrebbe apparire non troppo grave nella misura in cui sarebbe possibile porre in relazione le due operazioni (una reciproco dell'altra; ma in ogni caso già per desumere questa relazione sarebbe necessario il confronto di tutte le operazioni relative a una parola), si pensi a un caso come *coācla* per *cloāca*: qui le operazioni necessarie sarebbero:

²⁰ Nel testo che segue si usano per semplicità degli esempi latini, ma naturalmente esso vale per entrambe le lingue classiche.

```
@5*1=" "; @2*0="l"
```

ovvero una cancellazione ($l > zero$) e un'inserzione (l prima di o), apparentemente prive di qualsiasi rapporto e giustificazione linguistica. Ancora, si pensi al caso di *accersō* per *arcessō*: le operazioni potrebbero essere:

```
@2*1="r"; @5*1="s"
```

cioè la sostituzione di c con r e di r con s , linguisticamente del tutto ingiustificate se considerate da questa prospettiva formale e in isolamento.

Un simile stato dei dati implicherebbe quindi la necessità di complesse analisi a posteriori di tutte le operazioni di trasformazione per ogni parola si da individuare i loro rapporti, quando pure questo possa risultare fattibile dato che è facile aumentare la complessità del quadro con l'accumulo di più trasformazioni (cioè di più elementi *msp*). Lo scopo della strutturazione del testo qui proposta è invece quello di fornire un quadro il più possibile completo e di consentire una lettura automatica dei dati attingendovi in modo casuale (dunque non necessariamente sequenziale né completo), il che implica la necessità di mantenere il più possibile autonomo ogni suo elemento. In simili casi dunque, dove si ha a che fare con spostamenti (reciproci come per la metatesi in *leriquus* o singoli come nel caso di *coācla*), non sarebbe auspicabile definirli tramite cancellazioni, inserzioni e sostituzioni, oscurando la vera essenza del fenomeno alla base della grafia riscontrata. Il meccanismo di trasformazione dovrà quindi prevedere oltre a cancellazioni, inserzioni e sostituzioni anche spostamenti (*coācla* > *cloāca*) e inversioni (*leriquus* > *reliquus*); in tal modo sarà possibile rispettare la natura del fenomeno linguistico e per conseguenza renderla apparente sul piano formale. Senza entrare nei complicati dettagli di implementazione, il sistema consente di operare cancellazioni, inserzioni e spostamenti mantenendo sempre la forma originale come punto di riferimento, e a tale scopo l'elemento *msp* viene arricchito di ulteriori specificazioni relative al tipo di operazione effettuata:

- sequenza originale (at, run): la sequenza di *input*, definita da numero (@) ed estensione (*).
- tipo di operazione (op): assegnazione (=), spostamento (>), inversione (~).
- nuova sequenza (val): una sequenza letterale di caratteri oppure una sequenza tratta dalla parola originale e definita quindi anch'essa da numero (@) ed estensione (*).

Qualche esempio potrà illustrare meglio tali specificazioni:

- assegnazione semplice: *pena* > *poena*: @2*1="oe".
- inversione: *leriquus* > *reliquus*: @1*1~@3*1 (o anche @3*1~@1*1).
- spostamento semplice: *coācla* > *cloāca*: @5*1>@2*0.
- spostamento complesso: *accersō* > *arcessō*: @5*1>@2*0; @2*2="c"; @6*1="ss".

In quest'ultimo esempio @5*1>@2*0 indica lo spostamento della r , @3*1=" " ($cc > c$) la degeminazione della c (la cui geminazione nella forma originaria conserva la chiusura della sillaba precedente che altrimenti risulterebbe aperta dallo spostamento di r), e @6*1="ss" ($s > ss$) la geminazione della s (che parimenti mantiene il profilo quantitativo della parola rispetto alla forma originaria). Una simile descrizione formale consente pertanto non solo di generare la forma ortograficamente corretta (*arcessō*) a partire da quella originale (*accersō*), ma anche di definire con chiarezza assai maggiore la tipologia dei fenomeni linguistici alla base dell'errore (lo spostamento della r con i conseguenti aggiustamenti volti a mantenere l'originario profilo quantitativo della parola). Un procedimento automatico potrà dunque avvalersi di entrambi questi vantaggi per generare il testo nella sua forma corretta e un indice dei fenomeni linguistici e paleografici notevoli, entrambi ottenuti in modo del tutto automatico e a partire da un unico insieme di dati.

Il modello considerato sinora per la definizione del rapporto tra la forma originale e quella ortograficamente corretta soddisfa in misura ragionevole le esigenze di costruzione di indici automatici essenziali di tipo linguistico e paleografico, ma deve ancora essere valutato il suo aspetto tipografico, ovvero la possibilità di generare automaticamente a partire da esso una forma conforme alle convenzioni editoriali dei testi epigrafici: in altri termini, non solo occorre generare una forma come *poena* a partire da *pena*, ma anche dotarla di tutti i necessari diacritici editoriali per ottenere (sempre automaticamente) *p<o>ena*. Una prima implementazione di questa capacità rispetto agli algoritmi già illustrati per la generazione della forma ortograficamente corretta si basa sulla marcatura dei caratteri ed evita qualsiasi loro cancellazione. Essenzialmente ogni carattere può ricevere tre tipi di marcatura indicati per brevità con i seguenti simboli:

simbolo	marcatura
---------	-----------

=	carattere modificato
+	carattere aggiunto
-	carattere eliminato

Ad esempio, una forma come *coventionid* corretta in *conventionē*²¹ darebbe luogo alla marcatura:

c	o	n	v	e	n	t	i	o	n	e	d
		+							=	-	

Una volta ottenuta tale marcatura è facile passare alla notazione per diacritici $co\langle n \rangle vention^r e^1 \{d\}$, inserendo ogni sequenza marcata con + fra parentesi uncinata, ogni sequenza marcata con - fra parentesi graffe e ogni sequenza marcata con = fra mezze parentesi quadre. Si è visto che le operazioni di modifica effettuate dall'elemento m_{SP} riguardano assegnazione (*set*: una sequenza o zero sostituita a un'altra o zero), spostamento (*move*) o inversione (*swap*): occorre dunque definire il rapporto di ciascuna di esse con questo tipo di marcatura. Nel caso di spostamento o inversione la definizione è relativamente semplice:

- lo spostamento implica la cancellazione della sequenza dalla sua sede originaria (dunque marcatura con -) e la sua inserzione altrove (dunque marcatura con +): $co\tilde{a}cla > clo\tilde{a}ca$ ovvero $cl+oacl-a$, donde $c\langle l \rangle lo\tilde{a}c\{l\}a^{22}$.
- l'inversione implica la duplice modifica delle due sequenze invertite: $leriquus > reliquus$ ovvero $r=el=iqus$, donde $r^r e^r l^i quus$.

Naturalmente nel secondo caso (come in ogni caso in cui un carattere sia modificato, ovvero in ogni caso di marcatura =) l'edizione epigrafica dovrà aggiungere in apparato l'indicazione della forma originaria ("leriquus *lapis*").

Le difficoltà maggiori sorgono invece nel caso dell'operazione di assegnazione, che opera con sequenze di lunghezza variabile da zero a n , dove inoltre le due sequenze non hanno necessariamente la stessa lunghezza e possono essere composte da un qualsiasi numero di caratteri uguali o diversi: anche un caso banale come $cosul > consul$ infatti mostra come le sequenze originaria e corretta siano di lunghezza diversa (s contro ns) e la seconda includa un carattere uguale alla prima (in virtù delle considerazioni linguistiche riportate sopra non si tratta infatti di inserire una n dinanzi a s quanto di sostituire la s con il gruppo ns , sì da mantenere chiaro il rapporto linguistico tra le due forme). Anche un caso di correzione tanto semplice condurrebbe dunque a notevoli difficoltà per la marcatura, dato che occorrerebbe comparare le due sequenze in modo da riconoscere quanto in esse sia in comune e quanto differisca, senza tuttavia poter disporre di un punto di partenza comune: se infatti si iniziasse a confrontare le sequenze carattere per carattere (l'unico modo possibile per una macchina, le cui operazioni di base sono sempre atomiche) dall'inizio alla fine ci si troverebbe subito dinanzi a due caratteri diversi, s contro n di ns , per poi finire con zero rispetto a s . Una tale comparazione dovrebbe quindi concludere per la mancanza di qualsiasi somiglianza fra le due sequenze, il che è palesemente errato. Si potrebbe magari pensare di risolvere questo caso effettuando la comparazione a partire dalla fine delle sequenze, in modo da individuare la s comune a s come a ns , ma appare ovvio che al di là di questo banale esempio le situazioni di confronto possono risultare assai più complesse: basti pensare al citato rapporto fra *coventionid* e *conventionē*, dove una semplice comparazione carattere per carattere sia dall'inizio che dalla fine non condurrebbe comunque al risultato sperato.

Dinanzi alla complessità del problema del confronto fra due sequenze contenenti potenzialmente delle porzioni comuni occorre piuttosto adottare algoritmi più evoluti come il cosiddetto LMS (*Longest Matching Sequence*), il cui impiego più comune è probabilmente esemplificato dal confronto fra due testi (o due file di qualsiasi altro tipo) allo scopo di rilevarne le differenze, mentre un'applicazione importante riguarda la ricerca di somiglianze fra sequenze di DNA²³. Senza entrare nei complessi dettagli dell'algoritmo, è sufficiente osservare che esso consente di individuare tutte le porzioni in comune fra due sequenze di caratteri, sicché ad esempio nel caso del tutto fittizio di *conventionid* > *conventionē* l'algoritmo consente di isolare le porzioni in comune e definire come aggiunte, cancellazioni o modificazioni le differenze riscontrate, donde un risultato raffigurabile come:

²¹ Naturalmente l'esempio è del tutto fittizio ma comodo: una tale forma è effettivamente attestata (*sc. de Bacch.*) e appartiene al latino arcaico, con uscita analogica dei temi in -y e mancata notazione della n nel gruppo ns , sicché più che di correzione qui si tratta di 'traduzione' nella forma classica.

²² In realtà lo spostamento è complicato dalla variabilità della lunghezza della sequenza da sostituire, che può essere zero come nell'esempio di $co\tilde{a}cla > clo\tilde{a}ca$ ma anche maggiore di zero quando oltre a spostare un carattere esso debba essere sostituito ad altri esistenti. In ogni caso, un'adeguata implementazione tiene conto anche di queste complicazioni.

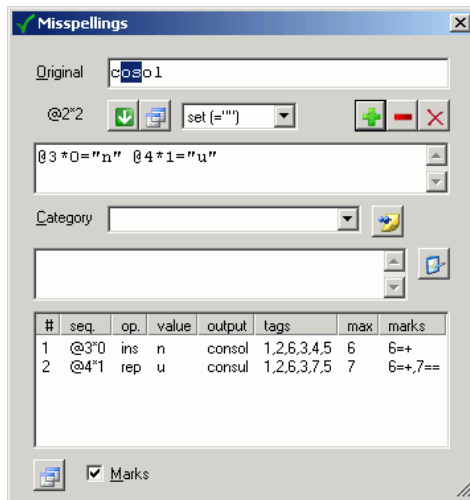
²³ Cf. ad es. Steven S. Skiena, *The Algorithm Design Manual*, New York 1997 (versione *online* all'indirizzo <http://www2.toki.or.id/book/AlgDesignManual/BOOK/BOOK/BOOK.HTM>).

```
co ventionid
conventionē
+      =-
```

(ovvero *n* è un'aggiunta, *e* una modifica di *i*, mentre l'originario *d* risulta cancellato). In tal modo, i complessi effetti di un'operazione di assegnazione di una sequenza a un'altra (entrambe di lunghezza variabile, donde possibili aggiunte, cancellazioni e ovviamente sostituzioni di caratteri) possono essere analizzati a posteriori si da consentire una marcatura basata sul confronto tra la forma originaria e quella trasformata²⁴.

Questo algoritmo associato alle altre funzionalità relative all'elemento *m*sp consente dunque di generare automaticamente una forma ortograficamente corretta e opportunamente marcata con i diacritici conformi alle tradizionali convenzioni epigrafiche: le aggiunte saranno inserite fra parentesi angolate, le cancellazioni fra graffe, e le correzioni fra mezze quadre (con relativa automatica aggiunta di una voce nell'apparato, ovvero nello strato dell'apparato), semplicemente risolvendo in sequenza ogni singolo elemento *m*sp proprio della parola in modo che l'*output* del precedente sia l'*input* del successivo sino alla fine della catena di trasformazioni, che condurrà alla forma corretta e marcata. Questo tipo di struttura consente così di utilizzare un'unica base di dati per attingervi di volta in volta le forme corrette delle parole con o senza una specifica marcatura di diacritici, o il materiale di base per la creazione di indici linguistici o paleografici, in entrambi i casi in modo del tutto automatico. Proprio l'aspetto dinamico di questi contenuti consente inoltre di variarne a piacere la forma, per adeguarsi ad esempio ad altre convenzioni editoriali o semplicemente modificare alcuni aspetti tipografici come colore, stile etc. più utili nell'ambito di pubblicazioni elettroniche.

Infine, si deve ancora una volta osservare che l'apporto di adeguati strumenti *software* non solo consente trasformazioni tanto complesse nella generazione dinamica dell'*output*, ma interviene anche in fase redazionale per facilitare l'immissione dei dati e in tal caso la definizione di tutti gli elementi *m*sp capaci di condurre dalla forma originale a quella ortograficamente corretta. A titolo di esempio si può riportare l'immagine di un'interfaccia grafica che consente al redattore l'immissione dei vari elementi *m*sp per una data parola del testo trattato²⁵:



interfaccia per l'immissione di elementi *m*sp

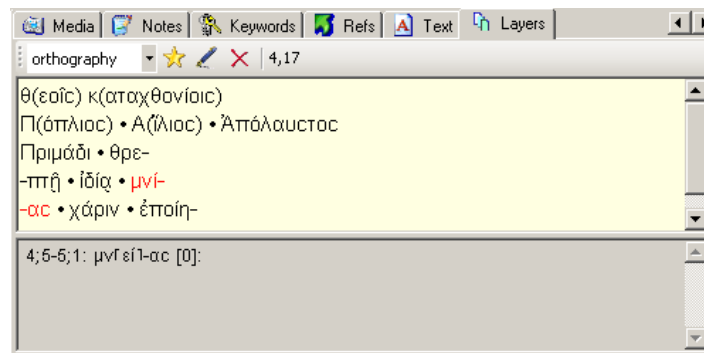
Questa finestra riporta in alto la parola come appare sulla pietra (qui *cosol*) e offre una serie di facilitazioni per la formulazione simbolica delle operazioni di trasformazione da effettuare per ottenere la grafia voluta (qui *consul*: inserzione di *n* e chiusura di *o* in *u*), elencate nel riquadro inferiore. Ad esempio, è sufficiente selezionare una porzione della parola per ottenere istantaneamente la sua corrispondente designazione simbolica (si noti qui la selezione di *os* in *consul* donde la relativa nota-

²⁴ Ci si potrebbe magari chiedere perché un tale algoritmo non sia adottato anche per automatizzare completamente la definizione del rapporto tra la forma della parola come appare sulla pietra e quella ortograficamente corretta: in realtà, a parte le limitazioni intrinseche all'algoritmo (per cui operazioni di spostamento e inversione produrrebbero spesso risultati scorretti), si è già visto come tale rapporto debba essere definito non solo in termini puramente meccanici ma secondo criteri il più possibile vicini al fenomeno linguistico che essi sottendono, sicché la decisione del redattore diviene comunque inevitabile.

²⁵ Come per l'interfaccia già illustrata a proposito della datazione, anche questo esempio è tratto dal programma redazionale per la generazione delle schede epigrafiche.

zione @2*2 sulla sinistra), e selezionare quindi una delle tre operazioni possibili (assegnazione, spostamento, inversione) perché la formula completa relativa alla trasformazione sia automaticamente inserita nella casella sottostante. In tal caso la casella contiene le due formulazioni per ottenere *consul* da *cosol*, @3*0="n" e @4*1="u", scritte di seguito; il programma si incarica di interpretare queste formule (che possono quindi essere accumulate dal redattore a formare lunghe espressioni), generare gli appositi elementi *m_{sp}* e inserirli nella catena trasformazionale. L'elenco delle trasformazioni viene aggiornato automaticamente a ogni nuova operazione, provocando il ricalcolo in tempo reale dell'intera catena di trasformazioni che conduce dalla forma originale a quella corretta: si offre così un immediato riscontro al redattore anche in relazione alle marcature utilizzate per l'inserzione dei diacritici nel testo pubblicato (qui si può notare una marcatura di aggiunta [+] per la *n* e una di modifica [=] per la *o*). Come si può osservare a ogni elemento *m_{sp}* può inoltre essere associata una categoria (utile per fornire una prima generica classificazione dei fenomeni linguistici) e annotazioni opzionali, entrambe definite per ogni *m_{sp}* nelle caselle centrali della finestra. Una volta immessi e verificati tutti gli elementi *m_{sp}*, il programma genera automaticamente il codice XML necessario alla loro inserzione nella banca dati.

Come si vede, per quanto possa essere complessa tutto questo tipo di informazione ortografica rimane strutturata e memorizzata a parte nel suo strato, in modo del tutto autonomo rispetto al testo di base da un lato e a qualsiasi altro strato ad esso associato dall'altro. Questo consente di realizzare quella architettura sempre aperta all'espansione che si è definita come una delle finalità dell'edizione qui presentata, e insieme di superare qualsiasi limitazione derivante dal basare l'edizione sulla semplice struttura portante del testo, che finirebbe per essere sovraccaricata di marcature sino a un livello di complessità ingestibile, oltre che incompatibile con la necessità di non alterare il materiale esistente e la sua struttura. La miglior conferma di questa espandibilità viene proprio dalla progressiva estensione del progetto al di là dei suoi confini originari per adattarsi di volta in volta alle nuove esigenze editoriali: a partire da un testo greco o latino con i soli strati ortografico e di commento (per qualsiasi commento generico associato a una porzione del testo originale) si sono aggiunti nel tempo una serie di nuovi strati per contenere dati relativi alle più diverse discipline dell'antichità, come filologia (strati di apparato critico per la costituzione del testo e strato delle citazioni letterarie in esso riportate), paleografia (strato riportante dati su nessi e legature del testo di base), storia (strato con dati prosopografici su ogni personaggio citato nel testo di base), ciascuno dotato della sua propria struttura e di una corrispondente interfaccia nel programma di immissione dei dati. In quest'ultimo un apposito pannello è deputato alla visualizzazione e immissione di informazione relativa a ogni strato: una volta selezionato lo strato di interesse da una casella a discesa il testo di base viene automaticamente colorato in modo da evidenziare tutta l'informazione associata in quello strato a ogni porzione di testo, consentendo la modifica o l'inserzione dei dati²⁶. Il redattore non deve quindi far altro che selezionare una porzione del testo di base e associarvi il tipo di informazione che desidera in base allo strato prescelto, senza alcuna nozione relativa alla struttura XML gestita dal programma.



particolare del pannello di modifica degli strati metatestuali: la porzione di testo cui sono associati dati nello strato selezionato è evidenziata in un colore diverso e il suo contenuto sinteticamente indicato in basso

Questi due esempi relativi alla codifica del testo greco o latino e alla sua strutturazione per righe e strati multipli, con tutte le annesse operazioni di trasformazione, illustrano a sufficienza sul piano meto-

²⁶ Come per numerosi altri aspetti del programma è facile immaginare la sua espandibilità, parallela a quella della banca dati da esso gestita, anche rispetto agli strati metatestuali: sarà infatti sufficiente aggiungere nuove voci alla casella degli strati e nuove finestre di dialogo per modificare i dati per rendere il programma in grado di gestire nuovi strati avvalendosi dell'interfaccia esistente.

dologico il tipo di implementazione adottato per i vari aspetti dell'edizione elettronica: unicità del contenuto e sua separazione dalle molteplici forme da esso rivestibili, scomposizione del contenuto in elementi atomici e loro strutturazione e classificazione semantica, trasformazioni automatiche più o meno complesse a ogni livello per la generazione di *output* derivanti dalla varia combinazione di questi elementi, il tutto realizzabile con un notevole apparato di strumenti *software* realizzati *ad hoc*. In tale ambito si comprendono meglio anche le prospettive filologica da un lato e informatica dall'altro alla base dei principi teorici del progetto, che tuttavia nelle intenzioni dell'autore possono progredire ben oltre anche sulla base della sua realizzazione.

Pubblicazione

Una volta terminata la fase di acquisizione e strutturazione del materiale, sia per i suoi aspetti multimediali che per quelli più strettamente testuali, una serie di file di immagine (fotografie e calchi) e file XML (banca dati delle iscrizioni e file di informazione meta-grafica) rigorosamente separati ma correlati fra loro costituisce il materiale completo della banca dati del *corpus*. A questo punto, ulteriori strumenti *software* sono deputati all'analisi e convalida del materiale e alla generazione automatica di tutte le risorse necessarie alla consultazione, anzitutto per la sua indicizzazione, avvalendosi di tecnologie più tradizionali come *database* relazionali per unire alla flessibilità di XML l'efficienza dell'SQL. Tutte le risorse necessarie sono comunemente generate automaticamente da appositi *software* di analisi e trasformazione, sicché l'intervento del redattore si limita qui all'utilizzo di tali strumenti.

A partire da questo stadio poi i percorsi redazionali possono diversificarsi in notevole misura a seconda del tipo di pubblicazione desiderato: *cd-rom* o *dvd*, sito *web*, chiosco multimediale, libro cartaceo o elettronico, etc. In generale, le funzioni di trasformazione di base saranno in comune fra tutti questi *output*, mentre per ognuno occorrerà poi creare del *software* specializzato in funzione della forma e del formato desiderati. Senza entrare nei complessi dettagli di simili funzioni, è però utile accennare a due sostanziali tipi di trasformazione di cui i programmi di pubblicazione sono capaci per evidenziare la versatilità dell'edizione anche nell'aspetto eminentemente testuale: si è già detto come la sua pubblicazione risulti del tutto indipendente da una specifica risorsa tipografica, dato che i convertitori euristici citati sopra sono utilizzabili per transcodificare il testo quale che sia il formato di *output* prescelto, consentendo così l'uso di qualsiasi *font* sia esso *Unicode* o meno. Ancor prima di questa operazione di transcodifica però si può collocare un'ulteriore serie di trasformazioni derivanti direttamente dalla 'competenza' che il programma ha del testo epigrafico inserito nella banca dati, guadagnata attraverso operazioni di *parsing* la cui complessità sfuggirebbe necessariamente a una tecnologia di trasformazione come XSLT. Il modo più semplice per illustrare in breve queste operazioni è citarne qualche esempio concreto, a partire dallo stesso programma di immissione dei dati, che include anche le funzioni di pubblicazione anzitutto a beneficio del redattore, abilitato a personalizzare ogni aspetto delle trasformazioni effettuate. Ad esempio si consideri la figura riportata qui sotto, relativa a una delle numerose interfacce offerte dal programma per queste operazioni di pretrattamento del testo:

- resolve abbreviations
- add dot after unresolved abbreviation
- keep type dd(omini)
- keep type C(aius)
- integrations
- litterae erasae
- litterae in litura
- additamenta
- litterae evanidae
- line above letters
- litterae Claudianae
- Greek diacritics
- Greek H=/h/ as breathing
- Latin diacritics
- Latin apex
- punctuation
- middle dot
- capitalize
- editorial spaces
- apparatus layer
- orthographical layer

Come si vede, è possibile scegliere se nel testo pubblicato si desidera sciogliere o meno le abbreviazioni (e nel secondo caso aggiungere un punto dopo le lettere relative all'abbreviazione non soluta), come trattare lo scioglimento di abbreviazioni particolari in rapporto al testo che ne deriva (ad es. quelle duplicate per indicare un plurale come *DD(omini)*, o quelle dotate di una lettera diversa nella forma abbreviata come *C(aius)*²⁷), se visualizzare o meno integrazioni editoriali, lettere *in rasura* o *in litura*, aggiunte del lapicida, lettere oggi scomparse dalla pietra, linee tracciate sopra le lettere nel testo originale; inoltre si può decidere se visualizzare lettere particolari come le claudiane o H = /h/ del greco o un sigma lunato, o sostituirle con le loro forme più standard (ad es. trasformando la H in spirito aspro o un sigma lunato nella sua forma standard, finale o meno a seconda del contesto), se visualizzare o meno tutti i diacritici moderni sulle lettere, come trattare quelli antichi come l'*apex*, se visualizzare l'interpunzione moderna e antica, se visualizzare tutte le lettere in maiuscola, se inserire o meno gli spazi che dividono le parole nell'edizione moderna ma in *scriptio continua* nel testo, etc. Inoltre, si può decidere il livello di interazione del testo di base con quegli strati metatestuali direttamente coinvolti nella sua trasformazione, come anzitutto l'apparato critico e lo strato ortografico: è infatti possibile sostituire automaticamente le grafie storiche con quelle normalizzate, inserendo gli opportuni diacritici editoriali (tipicamente parentesi), e le lezioni del testo originale con quelle proposte in apparato ed accettate dall'editore. Il programma è anche in grado di aggiungere automaticamente lemmi all'apparato critico quando una forma del testo sia normalizzata in base allo strato ortografico: se ad esempio l'utente sceglie di visualizzare il testo secondo l'ortografia standard e di visualizzare un apparato, il programma sostituirà autonomamente una forma come ἐπύησεν con ἐποίησεν nel testo originale, inserendo le parentesi indicanti la 'correzione' e aggiungendo in apparato un lemma come "ἐπύησεν lapis". In tal modo il *software* combina e riassume diversi tipi di informazione scomposti e memorizzati ciascuno nel suo contenitore, generando di volta in volta l'*output* richiesto dall'utente o dall'editore. Ad esempio, un testo del tutto fittizio come:

M(arcus) Iūl(ius) · Cae[sar], divus. μῆνιν ἄειδε, θεά, Πηληϊάδεω Ἀχιλῆος ἠόρπη

potrebbe essere pubblicato in forme diverse a seconda delle scelte effettuate:

<input type="checkbox"/> resolve abbreviations <input checked="" type="checkbox"/> add dot after unresolved abbreviation <input type="checkbox"/> keep type dd(omini) <input type="checkbox"/> keep type C(aius) <input type="checkbox"/> integrations <input checked="" type="checkbox"/> litterae erasae <input checked="" type="checkbox"/> litterae in litura <input checked="" type="checkbox"/> additamenta <input checked="" type="checkbox"/> litterae evanidae <input checked="" type="checkbox"/> line above letters <input checked="" type="checkbox"/> litterae Claudianae <input checked="" type="checkbox"/> Greek diacritics <input checked="" type="checkbox"/> Greek H=/h/ as breathing <input checked="" type="checkbox"/> Latin diacritics <input checked="" type="checkbox"/> Latin apex <input checked="" type="checkbox"/> punctuation <input checked="" type="checkbox"/> middle dot <input type="checkbox"/> capitalize <input checked="" type="checkbox"/> editorial spaces <input checked="" type="checkbox"/> apparatus layer <input checked="" type="checkbox"/> orthographical layer	<input type="checkbox"/> resolve abbreviations <input checked="" type="checkbox"/> add dot after unresolved abbreviation <input type="checkbox"/> keep type dd(omini) <input type="checkbox"/> keep type C(aius) <input type="checkbox"/> integrations <input checked="" type="checkbox"/> litterae erasae <input checked="" type="checkbox"/> litterae in litura <input checked="" type="checkbox"/> additamenta <input checked="" type="checkbox"/> litterae evanidae <input checked="" type="checkbox"/> line above letters <input checked="" type="checkbox"/> litterae Claudianae <input type="checkbox"/> Greek diacritics <input type="checkbox"/> Greek H=/h/ as breathing <input type="checkbox"/> Latin diacritics <input checked="" type="checkbox"/> Latin apex <input checked="" type="checkbox"/> punctuation <input checked="" type="checkbox"/> middle dot <input type="checkbox"/> capitalize <input checked="" type="checkbox"/> editorial spaces <input checked="" type="checkbox"/> apparatus layer <input checked="" type="checkbox"/> orthographical layer	<input type="checkbox"/> resolve abbreviations <input checked="" type="checkbox"/> add dot after unresolved abbreviation <input type="checkbox"/> keep type dd(omini) <input type="checkbox"/> keep type C(aius) <input type="checkbox"/> integrations <input checked="" type="checkbox"/> litterae erasae <input checked="" type="checkbox"/> litterae in litura <input checked="" type="checkbox"/> additamenta <input checked="" type="checkbox"/> litterae evanidae <input checked="" type="checkbox"/> line above letters <input checked="" type="checkbox"/> litterae Claudianae <input type="checkbox"/> Greek diacritics <input type="checkbox"/> Greek H=/h/ as breathing <input type="checkbox"/> Latin diacritics <input checked="" type="checkbox"/> Latin apex <input type="checkbox"/> punctuation <input checked="" type="checkbox"/> middle dot <input checked="" type="checkbox"/> capitalize <input type="checkbox"/> editorial spaces <input checked="" type="checkbox"/> apparatus layer <input checked="" type="checkbox"/> orthographical layer
M. Iūl. Cae, divus. μῆνιν ἄειδε, θεά, Πηληϊάδεω Ἀχιλῆος ἠόρπη	M. Iul. Cae, divus. μῆνιν αειδε, θεα, Πηληιαδεω Αχιληος ἠαορπη	MIULCAEDÍVUSMHNINAEI ΔΕΘΕΑΠΗΛΗΙΑΔΕΩΑΧΙΛΗ ΟΣΗΑΡΠΗ

²⁷ Simili opzioni sono finalizzate non solo a soddisfare diversi pubblici, ma anche a produrre un testo ulteriormente manipolabile in modo automatico senza che il programma si trovi dinanzi a *monstra* come *ddomini* o *Caius* invece dei corretti *domini* e *Gaius*.

Questo tipo di trasformazione dunque non rappresenta un semplice requisito tecnico, come nel caso della transcodifica testuale necessaria per generare qualsiasi codifica e formato digitali desiderati, ma consente all'editore come al fruitore dell'edizione di scegliere in ogni momento il tipo di informazioni da visualizzare nel testo, e quanto esso debba apparire vicino a quello originale, selezionando nel dettaglio ogni aspetto relativo alle convenzioni editoriali di un testo epigrafico. A tale scopo il programma si basa appunto sulla sua competenza del testo originale, di cui è in grado di individuare i diversi elementi come integrazioni, scioglimenti di abbreviazioni, interventi editoriali, uso di lettere e diacritici, etc. Peraltro, è importante osservare come tutti questi tipi di trasformazione non alterino in alcun modo il contenuto della banca dati: si tratta solo di definire diversi modi di attingere ad essi e di combinarli nel generare dinamicamente la forma desiderata per la pubblicazione. In tal modo si ottiene una grande versatilità dell'unico contenuto dell'edizione, capace di assumere forme virtualmente infinite in formati di volta in volta adatti al *medium* e al pubblico prescelti, proprio in virtù della semantizzazione e della scomposizione dei suoi dati effettuata dall'autore con l'ausilio del *software* di immissione dati. Inoltre, questa capacità di trasformazione può essere messa a frutto non solo per gli scopi della pubblicazione (per cui essa è ovviamente parte integrante del progetto editoriale), ma anche per convertire l'edizione in un qualsiasi altro formato (ad es. TEI), magari al fine di scambiare i dati con altri *corpora* digitali, o in prospettiva per adeguarsi al mutato panorama delle tecnologie informatiche mantenendo l'opera aggiornata e al riparo dal pericolo di una precoce obsolescenza.

Prospettive

Come si è ricordato, il progetto qui presentato è frutto di una costante evoluzione ed espansione, rese possibili dall'adozione sin dalla sua origine di una struttura appositamente concepita per essere il più possibile aperta e versatile, e rispetto alla quale si giustificano le varie scelte relative alle tecnologie adottate e la creazione (tuttora in corso) di tutta una serie di strumenti *software* capaci di operare in sinergia per supportare tutte le sue fasi editoriali. Se da un lato la sua recente espansione al mondo medievale, con tutte le necessarie aggiunte alla struttura della banca dati, testimonia concretamente di questa sua versatilità, anche l'inserimento di questo progetto editoriale nell'ambito di più ampie iniziative come quella promossa dall'Università di Sydney per la raccolta del materiale documentario relativo al teatro greco antico rappresenta l'occasione per dimostrare la funzionalità di questa architettura anche ben oltre gli scopi per i quali il progetto originario è stato avviato. Si è infatti visto che un aspetto fondamentale del progetto è costituito dall'uso di grandi e articolati contenitori XML che nelle prime fasi redazionali consentono di strutturare e memorizzare ogni tipo di informazione con grande duttilità e fluidità, requisito essenziale specie durante l'avvio di progetti editoriali di ampio respiro, per i quali è inevitabile partire con un semplice abbozzo di idea destinato a definirsi meglio nel corso stesso della raccolta e dell'analisi del materiale. Nel caso di questa espansione dunque tale caratteristica risulta ancora più utile nella misura in cui si tende a creare una grande banca dati dove l'aspetto epigrafico sarà in prospettiva solo uno dei numerosi considerati, accanto ad esempio a testi letterari e documenti materiali come vasi, ostraka, maschere teatrali, etc. In questo ambito sarà infatti possibile operare sempre all'interno della medesima struttura redazionale, e in certa misura con gli stessi strumenti *software* già disponibili, semplicemente affiancando all'«albero» dell'informazione relativo a testi epigrafici nuovi «alberi», ciascuno dei quali specializzato per contenere e strutturare al meglio un dato tipo di contenuti. La capacità di trasformazione implicita in ciascuno di essi farà poi sì che sia possibile unificare tutte queste fonti di informazione in un insieme coerente, opportunamente indicizzato e strutturato per la pubblicazione su diversi *media* e per diversi pubblici; nondimeno, le procedure generali e la filosofia alla base di questo progetto rimarranno inalterate, a dimostrazione della effettiva sua capacità di soddisfare gli scopi per i quali è stato concepito; un'unica banca dati altamente strutturata dal solo punto di vista semantico diviene così capace di trasformarsi in forme virtualmente infinite, adatte ai *media* e ai pubblici più disparati, e nel contempo di estendere i suoi contenuti nel senso della specializzazione come dell'espansione a dati persino privi di un contenuto testuale. Peraltro, il progetto qui illustrato è parte di un panorama che nelle intenzioni del suo autore può risultare ancora più ampio in virtù proprio delle prospettive filologiche e linguistiche cui accennavo in principio. Di fatto, la disponibilità di edizioni elettroniche di *corpora* testuali classici, al di là degli ovvi benefici come strumenti di studio e consultazione a ogni livello, da quello del visitatore occasionale fino allo studioso specialista, offre in prospettiva anche enormi potenzialità di analisi automatica. Molti degli strumenti *software* utilizzati in questo progetto infatti sono parte di un piano di lavoro assai più vasto, che trae origine dai miei interessi linguistici e metrici e si alimenta parallelamente di opere in corso di realizzazione per case editrici operanti nell'ambito delle lingue classiche. Una seconda generazione di *software* da me realizzato o in corso di realizzazione seguendo le ultime tecnologie nell'ambito dell'ingegneria del *software* per il mercato dell'informatica personale (anzitutto il *DotNet framework* di *Microsoft*) mira infatti a porre le basi per veri sistemi esperti delle

lingue classiche, finalizzati a ricerche concrete come ad esempio l'analisi automatica prosodica, metrica e linguistica di testi poetici²⁸, o la creazione di una grammatica storica elettronica della lingua latina e greca capace di offrire la completa flessione di ogni lemma del vocabolario completa di tutti i suoi stadi storici intermedi, ricostruiti o attestati. In questo senso, la disponibilità di *corpora* testuali di estensione paragonabile a quella degli attuali *cd-rom Packard Humanities* per la letteratura greca o latina, ma tratti da materiale epigrafico, di estremo interesse sul piano linguistico anzitutto per la preziosa testimonianza della viva evoluzione linguistica, potrebbe offrire interessanti campi di ricerca per analisi informatizzate a vasto raggio, capaci di fornire dati abbondanti, dettagliati e di prima mano su fenomeni metrici e linguistici di ogni sorta, raccolti con un metodo coerente e rigoroso nella misura in cui richiede una severa formalizzazione per essere compreso da una macchina. La disponibilità di materiale in formato elettronico può ovviamente costituire l'indispensabile premessa per simili ricerche, tanto più se come nel caso del progetto qui presentato i principi stessi alla base dell'edizione siano concepiti sin dall'inizio come funzionali ad applicazioni il più possibile diversificate e non strettamente epigrafiche, come indagini paleografiche o linguistiche (si pensi ad esempio al ruolo dell'elemento *msp* nella strutturazione del testo). Peraltro, a loro volta questi progetti possono fornire interessanti interazioni anche in rapporto all'edizione elettronica: ad esempio, al di là del rilevante e intrinseco interesse teorico, la creazione di un motore flessivo per una lingua classica offre poi numerose applicazioni pratiche specie nell'ambito dell'editoria elettronica, dalla quale d'altra parte un simile progetto trae origine (nell'immediato, il riconoscimento del lemma cui una forma flessa appartiene): per limitarsi solo ad alcuni esempi più da vicino connessi all'edizione epigrafica, la disponibilità di ogni forma flessa può essere sfruttata a scopi molto vari a seconda del contesto:

- in un *corpus* di testi, siano essi epigrafici, papiracei o 'letterari', diviene possibile lemmatizzare ogni singola parola riconducendo tutte le sue varie forme flesse all'unico lemma: è infatti evidente che consultando i prodotti della flessione forme come *laudābat*, *laudētis*, *laudāti erimus*, etc. potranno tutte essere ricondotte all'unico lemma *laudō*, fornendo anzitutto un'indicazione assai più significativa della frequenza testuale di ogni 'parola'. Peraltro, questo vantaggio non si limita all'aspetto 'statico' relativo ad esempio alla creazione di indici in un'edizione, ma sul piano 'dinamico' della pubblicazione elettronica una simile capacità rende infinitamente più efficace una ricerca all'interno di *corpora* testuali, tanto più in lingue flessive complesse come quelle classiche: in esse infatti una singola 'parola' può assumere anche centinaia di forme diverse a seconda della sua flessione, che sarebbe ben difficile far rientrare in espressioni di ricerca (senza peraltro considerare che la gran parte degli strumenti di ricerca offre capacità assai limitate per la loro formulazione, senza ricorrere neppure alla potenza delle espressioni regolari). Di fatto, attualmente si tenta di rimediare a questa situazione effettuando più ricerche per la stessa parola e limitandosi alle forme che si reputano più frequenti, oppure allargando le maglie della ricerca si da produrre un enorme numero di risultati non pertinenti. Avvalendosi invece della capacità flessiva del motore, sarebbe semplice richiedere la ricerca di tutte le forme di *laudō* affidando al programma di consultazione il compito di effettuare la ricerca richiesta in rapporto a tutte le possibili forme che tale parola può assumere nella sua flessione.

- un'ulteriore applicazione pratica di facile comprensione è in questo ambito una semplice specializzazione di quanto appena illustrato, in rapporto a edizioni elettroniche di natura epigrafica: si pensi ad esempio alla possibilità di costruire un repertorio completo delle abbreviazioni presenti in tutte le iscrizioni del *corpus*, dotato peraltro di un'informazione aggiuntiva che nessun repertorio cartaceo fornisce, come la frequenza testuale di ciascuna di esse. La creazione di un simile repertorio è possibile solo in virtù della capacità di lemmatizzare offerta dal motore flessivo, nella misura in cui ad esempio forme come *f(ilius)*, *f(ili)*, *f(iliū)*, *f(iliae)*, *f(iliūs)*, *f(iliārum)* etc. devono essere tutte ricondotte alla medesima parola cui corrisponde la stessa abbreviazione F, la cui frequenza sarà dunque costituita dalla somma delle frequenze di tutte le sue forme flesse.

- specie per testi spesso mutili come epigrafi o papiri, un repertorio di tutte le forme flesse a partire da un lessico sufficientemente completo della lingua latina (com'è quello che ha dato origine al progetto) potrebbe poi essere utilizzato come strumento ausiliario per l'editore stesso: si pensi ad esempio alla possibilità di suggerire tutte le possibili integrazioni di una parola mutila, di cui magari si conoscano solo alcune lettere iniziali e/o finali e un'estensione approssimativa: un programma potrebbe in tal caso effettuare una ricerca completa su tutto il repertorio delle forme flesse per

²⁸ Di un'analisi informatizzata relativa a testi epici greci per un totale di circa novantamila versi ho offerto un saggio in *Fra metrica e linguistica: per la contestualizzazione di alcune leggi esametriche* (in *L'esametro greco e latino: analisi, problemi e prospettive* a c. di E. Di Lorenzo, Atti del convegno di Fisciano 28-29 maggio 2002, pp.33-63.) e in *Appunti sulla prosodia del Lussorio di Shackleton-Bailey: alcune questioni di metodo*, in Luxoriana a cura di F. Bertini, pp.193-313, Genova 2002, frutto entrambi di una prima generazione di strumenti *software* realizzati in occasione della mia tesi di dottorato.

suggerire tutte quelle che formalmente risultino compatibili con lo spazio della lacuna. Naturalmente l'editore dovrebbe poi compiere comunque una scelta in base a criteri ben più rilevanti, ma lo spettro delle possibilità suggerite dal programma potrebbe in ogni caso costituire un comodo punto di partenza e considerare anche forme cui un editore potrebbe non giungere a pensare²⁹.

- un'applicazione banale di un tale repertorio può inoltre essere offerta da sistemi di controllo di testi elettronici, quale che sia la loro destinazione: per citare esempi ovvi in virtù della loro diffusione si possono citare i numerosi meccanismi di controllo ortografico implementati nei più comuni applicativi di videoscrittura o *desktop-publishing*, sicché ad esempio un tale repertorio potrebbe facilmente individuare inevitabili errori di digitazione presenti in *corpora* testuali o opere come vocabolari, siano essi destinati alla pubblicazione elettronica o cartacea.

- dal punto di vista strettamente linguistico, oltre all'ovvio interesse implicato dalla realizzazione di un simile motore, è facile immaginare le molteplici applicazioni cui sarebbe possibile piegare un repertorio di tutte le forme 'grammaticali' della lingua, ad esempio anche solo per l'analisi statistica delle strutture formali della lingua stessa (frequenze dei fonemi, ordine delle loro sequenze, tipi sillabici, configurazioni accentuali, configurazioni quantitative, etc.), con tutte le connessioni e le implicazioni a osservazioni generali di natura prosodica e metrica³⁰.

- rispetto ad altri strumenti informatici di analisi evoluta, come ad esempio quelli realizzati in concomitanza al motore per l'analisi prosodica e metrica di testi poetici classici, la capacità di lemmatizzare e individuare i rapporti sintattici fra parole all'interno del verso costituisce poi un interessante presupposto per l'ideazione di strumenti ancor più raffinati, capaci di riconoscere la strutturazione superficiale della frase si da aggiungere a criteri puramente formali informazioni di natura sintattica e semantica all'articolazione complessiva del verso.

Naturalmente questi sono solo alcuni esempi di applicazioni più o meno complesse, ma comunque in grado di mostrare le potenzialità scientifiche e pragmatiche derivanti non solo dalla progettazione di edizioni elettroniche concepite secondo principi teorici e metodologici aperti al più vasto campo di utilizzo, ma anche dall'interazione di diversi progetti accomunati dall'uso dello strumento informatico e per conseguenza dalla realizzazione di *software* estremamente specializzato: nell'ambito della programmazione fortemente orientata agli oggetti, la versatilità e il riutilizzo di ogni strumento creato all'interno di un progetto per contribuire alla realizzazione di altri progetti costituisce il paradigma di un'edizione capace non solo di presentarsi come strumento di ricerca o divulgazione con la massima espandibilità e versatilità possibili, ma anche fondamento di ulteriori progetti informatici rivolti ad altri campi delle scienze dell'antichità.

²⁹ Non sarebbe peraltro possibile obiettare che simili risultati si possano raggiungere anche tramite la mera indicizzazione di un vasto *corpus* testuale: è noto ad esempio che meritorie opere come il *Thesaurus Linguae Graecae* offrono un indice completo di tutte le forme che appaiono nei testi, che comunque risulta per sua natura diverso rispetto a quello offerto da un motore flessivo: laddove infatti il repertorio si limita a elencare le parole attestate nei testi inclusi nel *corpus*, senza peraltro consentire una loro lemmatizzazione, le forme flesse sulla base teorica da un motore possono naturalmente includere anche forme effettivamente non attestate nei testi in nostro possesso o inclusi nel *corpus*, ma del tutto legittime rispetto alla grammatica della lingua. Nessun *corpus* testuale infatti per quanto ampio e completo potrebbe contenere tutte le forme possibili della lingua latina o greca, incluse varianti dialettali o arcaiche: per fare un banale esempio di fantasia, il fatto che magari nessuno dei nostri testi includa una forma di genitivo come *rōsāī* per il termine *rōsa* non implica certo che una tale forma non sia esistita, ma semplicemente che essa, del tutto legittima rispetto alla grammatica latina, non è attestata nei testi a noi giunti. Dato che il motore flessivo genera tutte le forme 'grammaticali' di una parola essa potrà invece figurare nel suo repertorio di termini, e quindi essere adeguatamente considerata in tutti i casi in cui sia necessario effettuare ricostruzioni (come ad esempio per sanare lacune) o affrontare qualsiasi tipo di ricerca in cui quanto sia rilevante non sia l'effettiva attestazione ma la pura 'grammaticalità' di una forma.

³⁰ Si pensi ad esempio ai numerosi casi in cui sia necessario interpretare un fenomeno metrico per decidere della sua natura più o meno 'intenzionale' (fenomeno effettivamente ricercato dal poeta per scopi artistici) o 'accidentale' (fenomeno risultante semplicemente dall'interazione più o meno complessa di fattori di natura linguistica): un classico caso per il latino è ad esempio il rapporto fra configurazioni quantitative e accentuali nella clausola esametrica, donde l'apparente attenzione della versificazione latina alla regolazione di accenti in clausola che in realtà discende in larga parte dalla regolazione delle quantità (di cui l'accento latino è ovviamente funzione).