UN PROGETTO DI RICERCA

L'APPLICAZIONE DEL SOFTWARE SPI AI CODICI SENESI

Arianna Ciula – ciula@media.unisi.it
Dottorato in Scienze del Libro, Dipartimento teoria e documentazione delle tradizioni culturali, Università degli Studi di Siena

Il progetto di ricerca oggetto della presentazione tenuta all'*Intercultural workshop on Mediaval poetry and Digital resources* è compreso nell'area di studio che la comunità scientifica internazionale definisce *image based humanities computing area*[1]; in specifico, il campo di studio della tesi di dottorato attualmente in corso di elaborazione si colloca a metà tra le tradizionali discipline di studio del manoscritto e l'informatica umanistica. L'obiettivo primario della ricerca è creare uno strumento che supporti l'analisi paleografica sulla base della rappresentazione digitale delle scritture.

L'analisi morfologica della scrittura paleografica rientra in un insieme di criteri di studio del singolo manoscritto che comprende, tra gli altri, l'analisi letteraria e linguistica del contenuto del testo, le considerazioni codicologiche, lo studio dei segni abbreviativi e di interpunzione e, in generale, il contesto storico e interpretativo che ne favoriscono una lettura integrale. Una datazione dei manoscritti basata esclusivamente su criteri morfologici e stilistici delle scritture in esse contenuti è di per sé una pratica non fattibile e scientificamente scorretta.

L'uso dell'applicazione SPI (*System for Palaeographic Inspections*) che qui si presenta è concepito quindi all'interno di una più ampia metodologia che non può prescindere dall'analisi testuale del manoscritto e dall'analisi paleografica classica. Il programma, ideato presso il dipartimento di informatica dell'Università degli Studi di Pisa[2], è stato pensato infatti in qualità di supporto all'identificazione di somiglianze tra lettere appartenenti a manoscritti diversi vergati in scrittura libraria. L'intenzione del progetto è quella di mostrare come la rappresentazione digitale possa contribuire a descrivere sinteticamente una certa tipologia grafica e possa quindi rendere esplicito il confronto fra differenti stili scrittori.

1. IL CORPUS DEI MANOSCRITTI

Il progetto di ricerca della tesi di dottorato al secondo anno di svolgimento è stato focalizzato sui manoscritti in carolina dell'Italia centrale databili tra il X e il XII secolo. La ragione di una tale scelta è duplice: innanzitutto si costata una generale scarsità di studi paleografici specifici in merito e, in secondo luogo, nonostante l'eredità di documenti librari risalenti a tale

---

[1] In merito alla definizione della disciplina cfr. le discussioni nei due numeri del periodico «Computers and the Humanities», 36, s. 1 e 3, 2002.

[2] Gli sviluppi dell'applicazione sono stati presentati in varie tesi di laurea e di dottorato e in un articolo scientifico; cfr. DIEGO SONA, *Distanza Tangente per l'Analisi Morfologica in Campo Paleografico,* Tesi di Laurea, Università degli Studi di Pisa, aa. 1995 – 1996; G. IOMMI, *Segmentazione di Caratteri nelle Scrittura Librarie Medievali,* Tesi di Laurea, Università degli Studi di Pisa, aa.1996 – 1997; FABIO AIOLLI, *Sviluppo di un Sistema Basato sulla Distanza Tangente per l'Analisi Morfologica di Caratteri Paleografici,* Tesi di Laurea, Università degli Studi di Pisa, aa. 1998 – 1999; FABIO AIOLLI, M. SIMI, DIEGO SONA, A. SPERDUTI, ANTONINA STARITA, G. ZACCAGNINI, *SPI: a System for Palaeographical Inspections*, «AIIA Notizie», dicembre 1999 e ELISA BRUSCHI, *Un Sistema Avanzato per l'Analisi Morfologica di Documenti Librari,* Tesi di Laurea, Università degli Studi di Pisa, aa. 1999 – 2000.

intervallo cronologico sia di un ammontare piuttosto ingente, per quanto concerne le caratteristiche grafiche, siano esse di scuola, regionali o individuali, le caratteristiche testuali e i criteri di datazione, essa ha goduto di meno attenzioni e studi rispetto alle epoche precedenti. E' pertanto lacunosa la descrizione scientifica in termini di produzione e circolazione di testi; inoltre la letteratura in merito si è comprensibilmente concentrata sullo studio di centri limitati, giungendo a una conoscenza più specifica solo di alcune realtà circostanziate. In modo particolare il centro Italia, ad eccezione dell'area caratterizzata dalla tipizzazione *romanesca*, attende ancora un'analisi testuale approfondita da vari punti di vista, storico, filologico e paleografico.

Il corpus di codici prescelto è costituito da delle unità codicologiche conservate alla Biblioteca Comunale degli Intronati di Siena, la maggioranza delle quali proviene dalle antiche biblioteche dei vari conventi benedettini situati nel territorio intorno alla città toscana, conventi e monasteri il cui patrimonio è stato confiscato tra il XVIII e XIX secolo. Si tratta di codici dall'aspetto di lusso e dalle grandi dimensioni e di testi soprattutto di tipo liturgico. La varietà di prodotti grafici, la mobilità di copisti e miniatori insieme alla diffusione della cultura alfabetizzata impongono il contributo differenziato di discipline diverse così da permettere un'osservazione scientifica della grafia, un'analisi dettagliata dei valori decorativi del manoscritto, una conoscenza accurata del contenuto testuale. Gli studi passati di storia dell'arte e di paleografia – in special modo le opere del Garrison[3] e del Berg[4] - non sono stati in grado di individuare uno *scriptorium* comune ad una tale varia quantità e qualità di testi. A causa della disomogeneità dello stile scrittorio e delle rarissime informazioni sull'attività di copiatura nei conventi, i vari codici sono stati classificati come generalmente appartenenti all'area dell'Italia centrale.

L'osservazione e l'analisi dei codici in carolina del X-XII secolo conservati alla Biblioteca Comunale degli Intronati di Siena ha permesso di produrre delle descrizioni più o meno dettagliate delle caratteristiche morfologiche dei codici considerati. A seguito dell'analisi paleografica diretta e del reperimento bibliografico in merito allo studio della miniatura[5], le descrizioni sono state elaborate in modo da contenere i tradizionali campi di schedatura - numero di carte, datazione, dimensioni, provenienza, scuola di miniatura, mani, tipo di legatura. In questi fascicoli descrittivi, emerge già l'esigenza di mostrare, quanto più spesso è possibile, un raffronto grafico con l'immagine del manoscritto che renda visivamente esplicita la descrizione verbale.

All'interno dell'indefinito corpus librario - circa 40 codici, senza considerare i frammenti -, è possibile caratterizzare il *continuum* della scrittura, rilevando delle categorie estreme: dalla carolina non formalizzata del IX secolo all'estremo opposto della pregotica della fine dell'XII e *incipit* del XIII secolo. Sulla base della letteratura relativa ai codici senesi in esame[6], è possibile individuare inoltre almeno tre gruppi di codici e due estremi di riferimento:

Estremo carolina non formalizzata: IX-X secolo;

1° gruppo: XI secolo;

2° gruppo: XII prima metà;

3° gruppo: XII seconda metà;

Estremo gotica incipiente: XII ex. – XIII in.

Fuori da questa classificazione approssimativa rimane poi un gruppo di codici genericamente datati al XII secolo.

---

[3] Cfr. EDWARD B. GARRISON, *Studies in the history of mediaeval Italian painting*, I-IV, L' impronta, Firenze, 1984.

[4] Cfr. KNUT BERG, *Studies in Tuscan Twelfth Century Illumination*, Universitetforlaget, Oslo, 1968.

[5] In particolare Cfr. BENTE KLANGE ADDABBO, Il "Sermo Vulgaris" nella miniatura del secolo XI a Siena, in Scritti di storia dell'arte in onore di Roberto Salvini, Sansoni Editore, Firenze, 1984, pp.19-23, tavv.IV-V e ID., *Codici miniati della Biblioteca comunale degli Intronati di Siena*, Edisiena, Siena, 1987.

[6] L'unico riferimento catalografico esistente, non esaustivo e piuttosto sintetico, ma utilissimo per una prima indagine è L. AVITABILE, M.C. DI FRANCO, V. JEMOLO, *Censimento dei codici dei secoli X-XII*, «Studi medievali», anno XI, fasc.II (dicembre 1970), pp.1075-1101.

Questa situazione descrittiva piuttosto indefinita è stata utilizzata come caso studio per l'applicazione SPI, finora mai effettivamente testata su un corpus paleografico concreto.

2. IL SOFTWARE SPI

La scrittura, come ogni attività umana, non è uguale. Nessuna lettera si presenta per due volte scritta esattamente nello stesso identico modo; questo problema della variabilità complica ogni elaborazione automatica dell'immagine di una qualsiasi scrittura umana: si tratta di una concretizzazione di quello che, nella disciplina che studia il riconoscimento automatico delle forme, è chiamato problema dell'inconsistenza. Tuttavia, nell'ambito della paleografia, è ben noto che ogni variazione nella scrittura non è casuale, ma bensì affidata al controllo dell'intelligenza dello scriba che, avendo ben presente un modello calligrafico da perseguire, prevede l'aspetto che il testo dovrà assumere, evita le varianti incorrenti, in definitiva mira all'uniformità, in modo da produrre una versione coerente della calligrafia che intende realizzare.

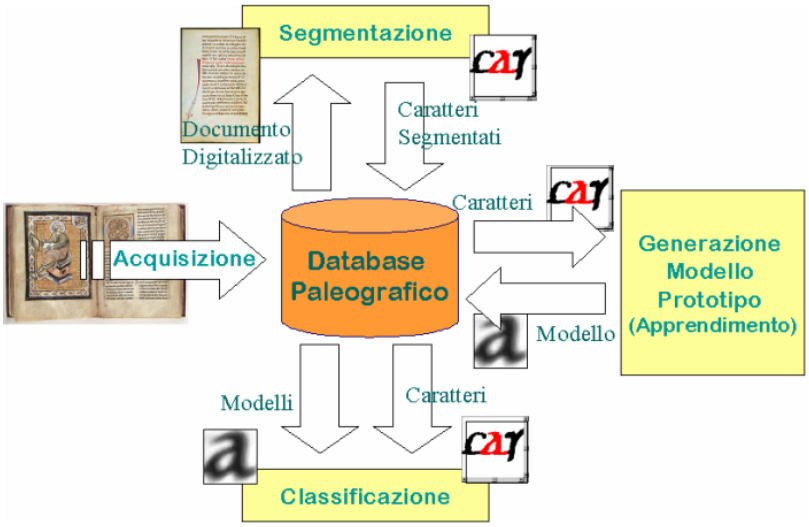
Anche nel caso in cui ci si serva di un sistema informatico quale l'applicazione SPI è dunque plausibile la scelta non riduttiva della singola lettera come base per l'analisi morfologica.

Oggetto dell'analisi paleografica classica sono, come è noto, forma, modulo, *ductus*, angolo di scrittura, tratteggio, legature, nessi e quante altre caratteristiche grafiche possano contribuire alla definizione di una tipologia scrittoria. L'applicazione SPI calcola ulteriori fattori relativi alla forma della scrittura ed è arricchita da parametri morfologici computabili sulla rappresentazione digitale e su di essa registrabili. I parametri computazionali così introdotti allargano il sistema classico di caratteristiche paleografiche propriamente dette con lo scopo ultimo di fornire degli indizi utili al paleografo per generare ipotesi sull'origine geografica e cronologica di manoscritti non datati e/o di provenienza incerta.

Le funzioni principali del sistema SPI implementato presso il dipartimento di informatica dell'Università di Pisa sono attualmente le seguenti:

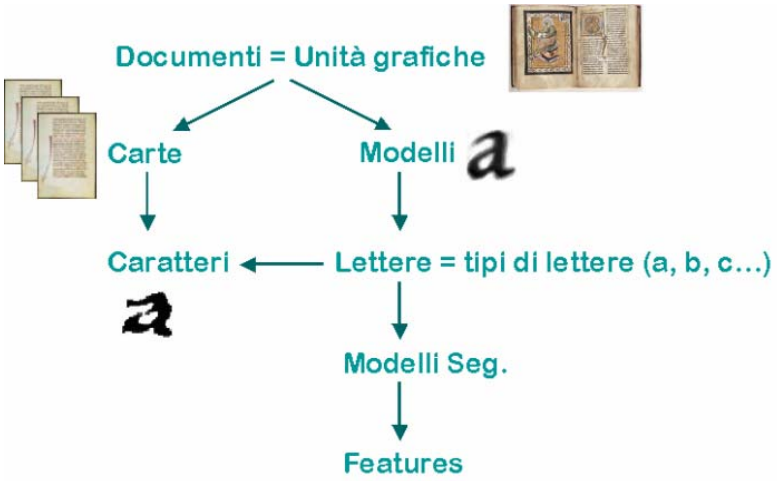
- mostrare e quantificare le relazioni grafiche tra manoscritti e stili scrittori diversi;
- fornire misurazioni di similarità oggettive tra documenti librari ignoti, per datazione e/o provenienza, e modelli caricati in memoria relativi ai manoscritti datati precedentemente analizzati.

Il sistema può essere visto come un insieme di moduli comunicanti mediante un database condiviso, ovvero un insieme di "agenti" che lavorano su moduli differenti: l'agente che si occupa della segmentazione, l'agente che gestisce la generazione dei modelli e l'agente che elabora l'interpretazione dei modelli.



Il sistema gestisce un database interno, codificato con il nome di database paleografico, atto a contenere tutta l'informazione necessaria nelle varie fasi. Il tipo di interazione richiesta all'utente si limita quindi all'inserimento nel database di un insieme di carte in formato bitmap - a seguito di un precedente lavoro di digitalizzazione dei manoscritti - e alla successiva estrazione guidata dal sistema dei caratteri d'interesse. I caratteri estratti, relativi ad una certa unità grafica, possono essere utilizzati a questo punto sia come esempi per la costruzione di un prototipo in un certo stila scrittorio (per esempio il modello della lettera *a* in una certa mano), sia come input al sistema di classificazione.

E' importante notare che la progettazione di un programma implica la definizione di una struttura del problema che si desidera gestire. Lo schema utilizzato per SPI è quello di un database relazionale, ovvero una struttura in cui ogni elemento incluso in un certo insieme si trova in relazione con un certo sottoinsieme di elementi appartenenti ad un altro insieme:



L'interfaccia grafica di SPI consta di una serie di schede, ognuna delle quali esegue una funzionalità specifica nel sistema. All'apertura del programma si accede ad un'unica scheda principale che rappresenta il punto di controllo di tutto l'applicativo.

Tramite la prima scheda del programma è possibile inserire nuove immagini bitmap di manoscritti digitalizzati; dalla stessa finestra si ha inoltre accesso al database paleografico. Esplorando il database, il paleografo può navigare fra le varie immagini di manoscritti librari inseriti nel sistema e le loro carte, visualizzare i caratteri estratti ed eventualmente cambiare lo stato di abilitazione per la fase di generazione dei modelli.

Nella sezione dedicata alle unità grafiche - impropriamente chiamate dagli informatici "documenti" - è possibile aggiungere nuove unità, cancellarne di esistenti o modificarne i dati corrispondenti. I dati di interesse per una certa unità manoscritta consistono nel nome dato dall'utente, che può coincidere con la segnatura del codice, la descrizione data dall'esperto e la datazione cronologica se conosciuta.

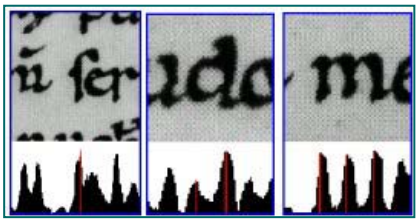
Nella sezione relativa alle carte è possibile ancora aggiungere o cancellare ogni singola carta. I dati rilevanti, in questo caso, consistono nel nome, che può coincidere con la numerazione effettiva della carta nel codice, e che descrive il cammino di ricerca della cartella in cui è contenuta l'immagine.

Per quanto riguarda la fase di segmentazione, la scelta effettuata è stata quella di non richiedere al sistema la segmentazione totale del documento, bensì delle singole lettere o gruppi di lettere connesse che risultano interessanti al singolo utente. E' da evidenziare che il modulo di segmentazione ha sinora riportato ottimi risultati di funzionamento.

La fase di segmentazione consta dei seguenti passi:

- all'interno del documento preventivamente scansito, il paleografo individua il tipo di lettera da segmentare per l'addestramento del sistema tramite la lista a discesa presente nella barra di controllo;
- a questo punto è necessario che venga individuata l'area in cui è contenuta la lettera, ovvero il *blob* che delimita l'effettiva estensione del carattere, con lo scopo di confinare la ricerca del carattere in una regione limitata dell'immagine[7];
- l'algoritmo di segmentazione segmenta la lettera o gruppi di lettere selezionate.

I parametri guida della segmentazione sono sintonizzati sulla tipologia del *pattern* da segmentare, ovvero su quella che viene chiamata proiezione *x* di un carattere[8]. Le lettere sono infatti idealmente divise in tre gruppi a seconda della distribuzione attesa delle aree di inchiostro[9]: caratteri spiccatamente 'unimodali' (ad esempio *i*, *l*) faranno parte del primo gruppo; quelli spiccatamente 'bimodali' del secondo (ad esempio *b*, la legatura *st*); quelli 'trimodali' del terzo (ad esempio *m*, la legatura *s*ti), mentre caratteri che hanno un istogramma atteso variabile a seconda dello stile scrittorio, come per esempio è il caso della lettera *g,* vengono di volta in volta segmentati selezionando la modalità corrispondente[10].



Per effettuare il passaggio alla seconda fase di generazione dei modelli, è necessario che la procedura di segmentazione venga ripetuta selezionando un certo insieme di caratteri con stesso valore semantico, vale a dire per esempio un certo numero di *a*.

La fase detta di apprendimento consta delle seguenti sezioni:

- dai caratteri estratti nella fase di segmentazione viene ricavato il *centroide*, ovvero una "lettera media", una sorta di "puro" rappresentante della classe, un modello appunto;

---

[7] Il blob ritagliato potrà contenere anche parti di altre lettere, senza per questo pregiudicare il successo delle fasi successive: è invece di capitale importanza che il blob racchiuda tutto il carattere da isolare, incluse le legature con le lettere connesse. L'individuazione del blob può avvenire o per selezione automatica, vale a dire che il sistema ricerca il box locale ottimale sulla base del puntamento del mouse approssimativamente nel centro della lettera (la procedura per la determinazione della box utilizza una tecnica indipendente dalla morfologia del carattere, basata esclusivamente sull'approssimazione locale delle dimensioni del carattere stesso, in base a criteri legati al tipo di scrittura: lo spessore del tratto, il modulo e il rapporto modulare fra altezza e larghezza della lettera), o per selezione manualmente esterna da parte del paleografo che seleziona con il mouse il blob d'interesse.

[8] Per proiezione di un carattere su una qualsiasi retta di pendenza *q*, si intende la somma degli elementi dell'immagine lungo una famiglia di rette perpendicolari a *q*; le proiezioni forniscono una buona indicazione della presenza di un oggetto in un'immagine, dove esso è localizzato e qual'é la sua estensione.

[9] Per misurare l'intensità luminosa dei pixel nell'immagine paleografica è utilizzato in specifico l'istogramma verticale.

[10] Nell'ultima versione di SPI viene offerta al paleografo la possibilità di consultare alcune variazioni della segmentazione minima ottenute mediante l'aggiunta di segmenti verticali alla segmentazione automatica di base, dando poi allo studioso la possibilità di optare per quella considerata migliore. Quindi da principio viene mostrata una segmentazione minima e a partire da quest'ultima viene offerta la possibilità di calcolare delle segmentazioni alternative che verranno proposte al paleografo congiuntamente alla soluzione minima e a quella consigliata dal sistema.

- i modelli così creati possono essere utilizzati come metri di paragone per classificare nuovi manoscritti digitalizzati o come oggetto di osservazione per un'analisi paleografica più approfondita.

Sulla base delle immagini digitali di documenti librari antichi manoscritti, il sistema genera quindi dei modelli stilistici relativi alle singole lettere o ai gruppi di lettere connesse. Tali prototipi costituiscono il tassello principale del sistema: essi infatti possono essere usati sia per l'interpretazione di manoscritti già memorizzati e quindi come oggetto di osservazione delle caratteristiche morfologiche per il paleografo, sia come classi di riferimento nella fase di classificazione di manoscritti mai presentati prima al sistema e quindi come modelli di riferimento per il sistema.

L'approccio presuppone quindi che ogni stile scrittorio sia rappresentato da un insieme di modelli di riferimento generati sulla base di un insieme di lettere dello stesso tipo. Ogni modello racchiude delle informazioni sulle caratteristiche principali della classe di lettere che rappresenta.

In termini matematici il prototipo è composto da un *centroide* più un insieme di tangenti di una certa cardinalità; il modello mediante le sue tangenti riesce ad approssimare i caratteri dell'insieme di apprendimento e rappresenta per questo una media speciale dei pattern per i quali è stato generato.

I modelli e in particolare la figura dimensionale che li rappresenta - il centroide - può subire una serie di trasformazioni definite tenendo conto della singola tipologia grafica esaminata. Questa procedura di trasformazione grafica è chiamata analisi delle tangenti e permette uno studio delle tangenti e quindi delle caratteristiche che il modello mette in evidenza rispetto al suo insieme di apprendimento. Il sistema non fa altro che visualizzare le proiezioni nel sottospazio del modello: vengono mantenute invariate le caratteristiche comuni o tangenti della classe, rendendo possibile invece lo spostamento del modello nel sottospazio generato dalle tangenti[11]. In questo modo, rendendo visibili le deformazioni del centroide lungo le tangenti relative è possibile osservare le direzioni di invarianza del modello stesso. Una finestra del programma come quella seguente permette appunto di visualizzare graficamente le tangenti e di modificare il prototipo del carattere variandone la posizione in un determinato sottoinsieme di tangenti.

.

In sostanza viene fornita la possibilità di 'stirare' in una direzione o in un'altra il centroide in modo da evidenziare le modifiche morfologiche che il modello permette. L'effetto visivo è un *morphing* del centroide. In questo modo è possibile ottenere informazioni utili sulle caratteristiche morfologiche principali del modello in esame, evidenziando al contempo le mutazioni morfologiche cui può essere soggetto un certo stile.

[11] In termini matematici, la varietà ottenuta al variare delle trasformazioni applicate al pattern viene approssimata con uno spazio lineare o spazio tangente di dimensione uguale al numero di trasformazioni definite.

Naturalmente è sempre possibile visualizzare e spostarsi tra i modelli generati. In particolare, l'indagine sull'insieme dei modelli inseriti è resa possibile dagli strumenti del dendogramma e del diagramma completo.

Il diagramma completo non è nient'altro che un grafico che rappresenta le relazioni tra documenti immagazzinati in memoria calcolando le distanze tra modelli di diversi stili scrittori. Per ogni prototipo, è possibile visualizzare il resto dei prototipi ordinati in base alla distanza dal modello in esame. Il diagramma completo presenta quindi per ogni modello del sistema l'insieme degli altri prototipi ordinati per grado di similarità.

È necessaria una fase preliminare nella quale l'applicazione calcola la matrice delle distanze fra i modelli del sistema presi a due a due. Ad ogni modello viene quindi associato un grado di similitudine rispetto a tutti gli altri. Questa stessa fase preliminare è necessaria anche per il calcolo del dendogramma.

La modalità di visualizzazione dell'insieme di modelli a dendogramma è basata su un algoritmo di *clusterizzazione* che suddivide l'insieme di modelli in sottogruppi secondo una modalità gerarchica. Ogni gruppo rappresenta un sottoinsieme di modelli accomunati da una certa similitudine morfologica. Il risultato di questo tipo di computazione può essere visualizzato mediante un albero binario chiamato appunto dendogramma. Tale schema associa ad ogni sottoalbero un certo *cluster* che contiene tutti i nodi foglia del sottoalbero.

L'elaborazione di un tale albero binario, di cui vedremo qualche esemplificazione sui codici senesi, rivela dunque le similarità tra gruppi di modelli di lettere, fermo restando che per ogni classe di lettere viene rappresentato il solo centroide.

I modelli, oltre che essere utili all'analisi paleografica e al confronto grafico, sono il riferimento per l'ultima fase progettata, vale a dire per la classificazione *tout court*. Per classificazione di un manoscritto si intende in questo caso il recupero dei modelli contenuti nel database del sistema che risultino più simili, da un punto di vista morfologico, allo stile scrittorio che si intende classificare o definire.

Nella fase operativa di classificazione di documenti ignoti, il paleografo estrae nuovamente delle lettere-campione dal manoscritto ignoto, ma questa volta non per addestrare il sistema, bensì per testarne la capacità classificatoria:

- ai campioni vengono applicate delle trasformazioni prevedendo le varianti possibili del carattere "puro".

;

- il sistema confronta i caratteri estratti dai nuovi documenti con i prototipi di riferimento memorizzati, appartenenti ai diversi documenti su cui era stato effettuato l'apprendimento – si deve trattare ovviamente di modelli abilitati nel sistema che siano compatibili con il carattere in esame (stesso tipo di lettera, stesso formato etc.) - : viene calcolata la cosiddetta distanza-tangente tra un carattere estratto - in tutte le sue occorrenze - e uno dei modelli generati relativi a un specifico manoscritto o stile scrittorio definite nei termini di distanza-tangente[12];

[12] In sostanza viene calcolata la distanza invariante alle trasformazioni tra i sottospazi delle rappresentazioni digitali prese a due a due. La distanza-tangente ha la proprietà di essere invariante rispetto a trasformazioni globali non

- il sistema individua il manoscritto i cui modelli-carattere risultano essere i più simili, dal punto di vista morfologico, ai nuovi caratteri estratti.

Questo tipo di classificazione è detta per comitato. Alla fine del procedimento, ovvero dopo aver processato l'insieme di caratteri-campione, la classificazione 'elegge' infatti lo stile scrittorio che abbia raggiunto il massimo di confronti vincenti.

Riassumendo, Il sistema supporta quindi:

- l'analisi morfologica del modello relativo a una singola lettera o gruppo di lettere;
- l'analisi comparata di un insieme preventivamente definito di modelli diversi;
- la classificazione di nuovi manoscritti mediante un insieme di classificatori che lavorano sulle singole lettere o gruppi di lettere connesse.

L'interfaccia grafica del programma, nonostante sia stata progettata in modo da garantire una gestione adeguata dell'intero database, costituito da immagini di manoscritti, caratteri estratti e prototipi o modelli generati, è stata sviluppata nella prospettiva di un analisi spiccatamente informatica e necessita dunque dei miglioramenti in vista di una utilizzazione in ambito prettamente paleografico e di una generale facilitazione dell'accesso.

3. SPERIMENTAZIONI SUL CORPUS SENESE

Il processo di uso e test dell'applicazione pisana è stato improntato su una relazione il più possibile biunivoca tra i diversi contributi disciplinari, informatico l'uno e paleografico l'altro. L'obiettivo primario e conclusivo dell'intero progetto è quello di identificare alcune caratteristiche discriminanti tra categorie scrittorie cronologicamente e geograficamente vicine, tutte appartenenti alla grande classe della carolina e all'interno di essa variamente distribuite. In concreto, questa sorta di descrizione particolareggiata delle varianti morfologiche della carolina tra X e XII secolo nella Toscana centrale è stata realizzata servendosi sia della consultazione della letteratura paleografica classica – che abbiamo preannunciato essere di consistenza piuttosto esigua -, sia di ricerche alternative, quali per esempio l'analisi dell'interpunzione e della decorazione. All'osservazione dei modelli ottenuti mediante l'applicazione pisana è pertanto attribuito lo stesso valore conoscitivo di una metodologia di analisi alternativa.

Per ogni codice del corpus è stata effettuata la scansione di almeno quattro carte nel caso in cui lo stile scrittorio sembrasse rispondere ad una certa uniformità o di più carte nel caso in cui occorressero più mani visibilmente disomogenee, mantenendo la proporzione di quattro carte per stile. La risoluzione prescelta per le immagini è 300 dpi, mentre il formato è stato convertito in *bitmap* con quantizzazione a due livelli come richiesto per l'elaborazione tramite il software SPI.

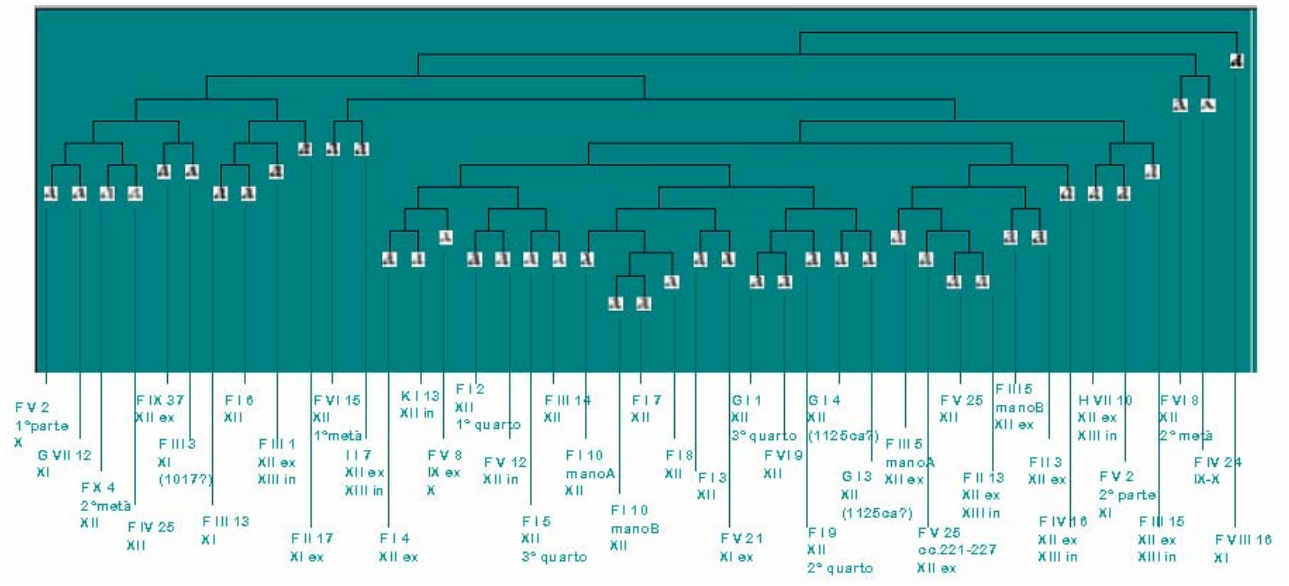
Per facilitare la gestione delle immagini nel database le carte sono state ritagliate e introdotte a sezioni – per colonna quando possibile. Per ogni carta è stata poi effettuata la segmentazione cominciando dalla lettera la cui evoluzione risulta piuttosto emblematica nello sviluppo della scrittura carolina: la lettera *a*.

Ci limiteremo in questa sede a dare una breve esemplificazione della generazione dei modelli e delle relazioni grafiche individuate in merito appunto alla sola lettera *a*.

Il dendogramma risultante dal raffronto automatico tra i diversi modelli ha prodotto un albero ramificato piuttosto complesso. Ad ogni prototipo è stata associata manualmente la segnatura del codice senese corrispondente:

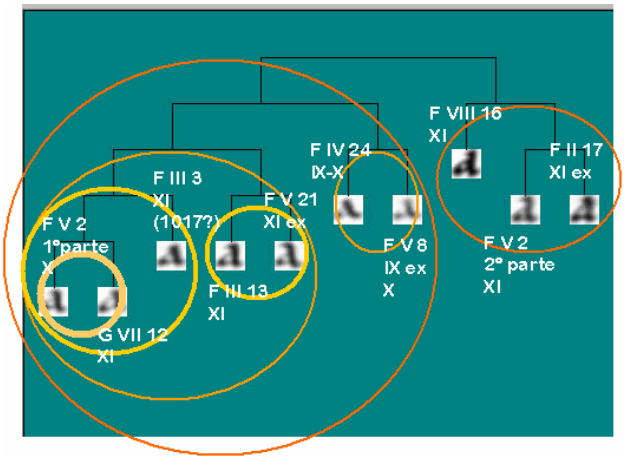
---

lineari, ovvero rispetto a tutte quelle distorsioni spesso involontariamente introdotte in fase di acquisizione, come traslazioni e rotazioni e di catturare, mediante procedure di apprendimento, le caratteristiche morfologiche rilevanti che identificano lo stile scrittorio di un particolare manoscritto.



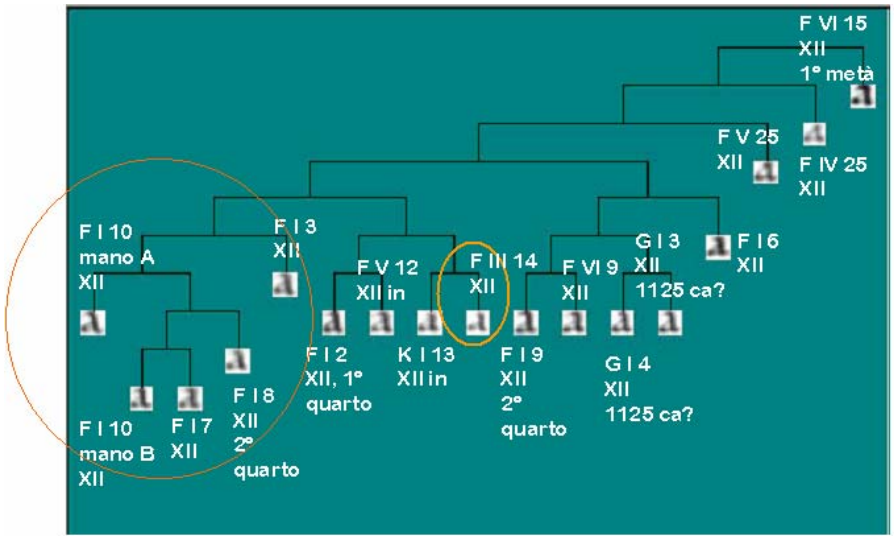
Il fatto che sia possibile memorizzare all'interno del database paleografico le immagini digitali, le notazioni storiche ad esse di volta in volta associate, le lettere segmentate e i modelli relativi costituisce già un ausilio non indifferente per l'accesso ai dati grafici e per lo studio sincronico dei manoscritti. Tuttavia a noi interessa esplorare la validità del supporto informatico nella sua relazione diretta con l'analisi paleografica e con l'interpretazione sulla datazione e geografia delle scritture. Per evidenziare questa funzionalità, è più utile analizzare l'elaborazione del dendogramma per intervalli cronologici distinti.

Considerando per esempio l'intervallo cronologico che comprende i secoli IX e X, la distribuzione delle somiglianze grafiche per i vari modelli relativi alla lettera *a* risulta la seguente:



L'intervallo temporale di un secolo genericamente indicato nella catalogazione apparsa su Studi Medievali relativamente ai codici senesi, si presenta arricchita di raggruppamenti di modelli proposti dal sistema e da noi evidenziati con l'uso dei cerchi colorati (a colore diverso corrisponde un diverso grado di ramificazione). L'intervallo di due secoli si ramifica in due sottogruppi, il primo dei quali a sinistra si suddivide a sua volta in due gruppi più ristretti, il primo dei quali individua ancora altri due rami, di cui sempre il primo presenta ancora l'ultima ramificazione: in tutto 7 insiemi che, ad un'analisi integrale di tutti gli elementi di studio del manoscritto, potrebbero supportare l'ipotesi di vicinanze di composizione e di stili scrittori.

Immaginiamo ora di abilitare soltanto quei modelli di lettere *a* relativi soltanto ai codici genericamente datati al XII secolo o alla prima metà del XII ancora dal catalogo del 1970 e di avviare l'elaborazione del dendogramma.



Concentrando ora l'attenzione sul gruppo cerchiato a sinistra notiamo che in esso è presente il modello della lettera *a* ricavato dalle carte manoscritte fotografate dal codice con segnatura F I 8, codice che nel catalogo viene datato al 2° quarto del XII secolo. E' possibile restringere i margini di datazione anche per il resto di codici raggruppati nel medesimo ramo del dendogramma e quindi ritenuti dal sistema graficamente simili per quanto concerne il modello in questione?

Ebbene uno studio della bibliografia dei manoscritti F I 3 e F I 7 permette per esempio di avanzare effettivamente una tale ipotesi di datazione. Entrambi i codici citati presentano una decorazione che è stata attribuita nel caso di F I 3 alle seconda metà del XII secolo e nel caso di F I 7 proprio al secondo quarto dello stesso secolo.

Il secondo cerchio più piccolo evidenzia il modello di *a* del codice con segnatura F III 14, genericamente datato nel catalogo al XII secolo. SPI ha inserito questo modello in un sottogruppo costituito tutto da codici del XII secolo *ineunte*: anche in questo caso l'evidente, ma parziale risultato dell'applicazione raffrontato con altri elementi di analisi potrebbe condurre alla revisione della datazione di F III 14. Si tratta ovviamente di inferenze suggerite che necessitano di uno studio approfondito prima di essere proposte come risolutive di una datazione generica; è tuttavia tramite tali rappresentazioni grafiche – in principio tramite la casella di analisi delle tangenti e quindi tramite gli strumenti del diagramma completo e del dendogramma - che il sistema può insinuare e supportare delle ipotesi paleografiche.

Essendo il lavoro di analisi paleografica e automatica dei codici senesi tuttora in fieri, rimandiamo esemplificazioni più esaustive alla tesi di dottorato in corso di elaborazione.

E' da anticipare inoltre che, a seguito della formalizzazione di ulteriori conoscenze paleografiche e storiche ritenute rilevanti e pertinenti rispetto al corpus paleografico scelto come oggetto di analisi, sono stati previsti ampliamenti e miglioramenti che ci auguriamo di poter apportare in futuro al sistema pisano.

BIBLIOGRAFIA

LA SCRITTURA CAROLINA SECC. X-XII

B.BISCHOFF, *Centri scrittori e manoscritti mediatori di civiltà dal VI secolo all'età di Carlomagno*, in G. Cavallo (ed.), *Libri e lettori nel medioevo. Guida storica e critica*, Laterza, Roma-Bari, 1977, pp. 29-72.

G. CENCETTI, *Postilla nuova a un problema paleografico vecchio: l'origine della minuscola "carolina"*, in «Nova Historia», VII (1955), pp. 9-32.

ID., *Scriptoria e Scritture nel Monachesimo*, in Atti de Settimane di Studio del Centro italiano di studi sull'Alto Medioevo, *Il Monachesimo nell'Alto medioevo e la formazione della civiltà occidentale*, IV (Spoleto 8-14 aprile 1956), Spoleto, 1957, pp. 188- 219.

ID., *Scritture e circolazione libraria nei monasteri benedettini*, in G. Cavallo (ed.), *Libri e lettori nel medioevo. Guida storica e critica*, Laterza, Roma-Bari, 1977, pp. 75-97.

MONIQUE-CECILE GARAND, *Manuscrits monastiques et scriptoria aux XIe at XIIe siècles*, in *Codicologica,* 3*, Essays typologiques*, E. J. Brill, Leiden, 1980.

MICHAEL GORMAN, *Bernard Bishoff Handlist of Carolingian manuscripts*, in «Scrittura e Civiltà», XXV, 2001, pp.89-112.

ARMANDO PETRUCCI, *Istruzioni per la datazione*, premesse a AA.VV., *Censimento dei codici dei secoli X-XII*, in «Studi medievali», 3^a s., IX/2, 1968, pp.1115-1126.

ALESSANDRO PRATESI*, Le ambizioni di una cultura unitaria: la riforma della scrittura*, in *Nascita dell'Europa carolingia: un'equazione da verificare*, Spoleto, 1981, pp. 507-523.

LUIGI SCHIAPARELLI, *Note paleografiche. A proposito di un recente articolo sull'origine della minuscola carolina*, in «Archivio storico italiano», 7° s. (1926), pp.3-23.

F. STEFFENS, *Paléographie latine : 125 Fac-similés en phototypie accompagnés de transcriptions et d' explications avec un exposé systématique de l' histoire de l' écriture latine*, Multigrafica, Roma, 1982.

CODICI DELLA BIBLIOTECA DEGLI INTRONATI DI SIENA

L. AVITABILE, M.C. DI FRANCO, V. JEMOLO*, Censimento dei codici dei secoli X-XII*, «Studi medievali», anno XI, fasc.II (dicembre 1970), pp.1075-1101.

KNUT BERG, *Studies in Tuscan Twelfth Century Illumination*, Universitetforlaget, Oslo, 1968.

G. M. CAO, T. CATALLO, M. CURANDAI, E. DI MATTIA, P. E. FORNACIARI, E. PERUZZI, F. SANTI (a cura di), *Catalogo dei manoscritti filosofici nelle biblioteche italiane*, VIII, Olschki, Firenze, 1996, pp. 101- 134.

LORENZO ILARI, *Indice per materie della Biblioteca comunale di Siena*, Tipografia All'insegna dell'ancora, Siena , 1844-1848.

Inventario topografico dei manoscritti della Biblioteca Comunale di Siena, BCS, ms. non inventariato, sec. XIX.

EDWARD B. GARRISON, *Studies in the history of mediaeval Italian painting*, I-IV, L' impronta, Firenze, 1984.

BENTE KLANGE ADDABBO, *Fogli Senesi inediti dei sec. XI e XII*, in *La Miniatura Italiana in età romanica e gotica.* Atti del I Congresso di Storia della miniatura italiana (Cortona 26-28 maggio 1978), Olschki, 1979, Firenze, pp. 103-117.

BENTE KLANGE ADDABBO, *Il "Sermo Vulgaris" nella miniatura del secolo XI a Siena*, in *Scritti di storia dell'arte in onore di Roberto Salvini*, Sansoni Editore, Firenze, 1984, pp.19-23, tavv.IV-V.

BENTE KLANGE ADDABBO*, Codici miniati della Biblioteca comunale degli Intronati di Siena*, Edisiena, Siena, 1987.

PALEOGRAFIA E INFORMATICA

ANDREA BOZZI, *Better Access to Manuscripts and Browsing of Images*, CLUEB, Bologna, 1997.

ANDREA BOZZI (ed), *Computer-aided recovery and analysis of damaged text documents*, CLUEB, Bologna, 2000.

SYLVIE CALABRETTO, ANDREA BOZZI, JEAN MARIE PINON, *Numérisation des manuscrits médiévaux: le projet européen Bambi*, in Atti del convegno *Vers une nouvelle érudition: numérisation et recherche en histoire du livre*, Rencontres Jacques Cartier, Lione, dicembre 1999.

ALESSANDRO PRATESI, *A proposito di tecniche di laboratorio e storia della scrittura*, in «Scrittura e civiltà», I (1977), pp.199-209.

Les tecniques de laboratoire dans l'étude des manuscrits, *Colloques Internationaux du CNRS*, n° 548, Parigi, 13-15 settembre 1972, Parigi, 1974.

«Computers and the Humanities»: 36, s. 1 e 3, 2002.

Il software per l'analisi paleografica SPI

FABIO AIOLLI, *Sviluppo di un Sistema Basato sulla Distanza Tangente per l'Analisi Morfologica di Caratteri Paleografici,* Tesi di Laurea, Università degli Studi di Pisa, aa. 1998 – 1999.

FABIO AIOLLI, M. SIMI, DIEGO SONA, A. SPERDUTI, ANTONINA STARITA, G. ZACCAGNINI, SPI: a System for Palaeographical Inspections, in «AIIA Notizie», dicembre 1999.

ELISA BRUSCHI, *Un Sistema Avanzato per l'Analisi Morfologica di Documenti Librari,* Tesi di Laurea, Università degli Studi di Pisa, aa. 1999 – 2000.

G. IOMMI, *Segmentazione di Caratteri nelle Scrittura Librarie Medievali,* Tesi di Laurea, Università degli Studi di Pisa, aa.1996 – 1997.

DIEGO SONA, *Distanza Tangente per l'Analisi Morfologica in Campo Paleografico,* Tesi di Laurea, Università degli Studi di Pisa, aa. 1995 – 1996.