Claude Cazalé Bérard Université Paris Ouest Nanterre La Défense

Critica genetica, Humanities Computing e Digital Humanities (Critique Génétique et Humanités numériques)

Premessa
1. Presentazione dell'ITEM (Institut des Textes et Manuscrits Modernes)
2. L'Equipe Humanités numériques
3. La base dati <i>MUSE</i> (Manuscrits, Usage des supporte et de l'Écriture)
4. Modalità e programmi di trascrizione per le edizioni elettroniche dei manoscritti.

Critica genetica e Humanities Computing (Critique Génétique et Humanités numériques)

Premessa

L'ormai pluridecennale esperienza acquisita da numerosi gruppi e istituti di ricerca universitari nel campo dell'Informatica Umanistica (*Humanities computing*) e nell'ambito, in pieno sviluppo, dell'archivistica digitalizzata (*Digital Humanities*), nelle più grandi biblioteche in Europa e negli Sati Uniti, non ha più da dimostrare la propria validità scientifica.

Molti sono i progetti che si sono sviluppati e hanno raggiunto un livello notevole di realizzazione in Francia a cominciare dal campo della medievistica, nell'archivistica, nell'archeologia, nelle biblioteche, ma è sicuramente nell'ambito dell'ITEM (Institut des Textes et Manuscrits modernes, ENS-CNRS) che è stata affrontata con più consapevolezza la rivoluzione epistemologica dovuta all'incontro tra discipline umanistiche (Sciences Humaines) e informatica, poiché si sono sviluppate di pari passo una ricerca teorica di stampo genetico e un'attività critica applicata all'edizione dei testi in tutte le loro fasi di elaborazione. Si suggerisce per tanto di consultare il sito (http://www.item.ens.fr) particolarmente ricco per l'abbondante documentazione, nonché per i numerosi progetti, programmi in corso e attività annunciate delle varie Équipes.

1. Presentazione dell'ITEM (Institut des Textes et Manuscrits Modernes)

L'ITEM è un Laboratorio di ricerche che dipende istituzionalmente dall'École Normale Supérieure e dal CNRS. Vi lavorano diverse Équipes che sviluppano progetti riguardanti manoscritti (francesi, francofoni, anglofoni, latino-americani), generi (autobiografia e corrispondenza, arti visive, fotografia, cinema), autori della modernità letteraria (Flaubert, Zola, Proust, Joyce, Aragon, Barthes, Celan) e filosofica (Illuminismo, Nietzsche, Valery, Sartre), linguistica (multiliguismo, traduzione...), informatica umanistica.

Da tempo l'incontro fra informatica e discipline umanistiche ha suscitato ricerche e progetti in seno all'ITEM, in particolare all'iniziativa di Jean-Louis Lebrave e Daniel Ferrer che tra i primi hanno proposto un'apertura della critica genetica alle nuove tecnologie, con riflessioni

teoriche e sperimentazione di ipertesti per l'edizione elettronica dei manoscritti. In questa prospettiva è stata creata, nel 2012, l'Équipe Humanités numériques.

2. L'Équipe Humanités numériques

L'Équipe è affidata ad un Collettivo internazionale:

Thomas Bartscherer (Ricercatore associato, New York)

Lou Burnard (Consulente)

Auerèle Crasson (Direttrice associata dell'ITEM)

Paolo D'Iorio (Direttore)

Nathalie Ferrand (Direttore di Ricerca)

Daniel Ferrer (Direttore di Ricerca)

Serge Linkes (Professore associato)

Jérémy Pedrazzi (Ingegnere informatico)

Armin Schweher (Ingegnere)

Richard Walter (Ingegnere di Ricerca)

L' Équipe ha messo a disposizione dei ricercatori e degli studiosi uno spazio per discussioni teoriche (convegni, seminari...) e per la sperimentazione dei progetti in corso. Il suo compito è anche di riunire e dare accesso a tutti i lavori compiuti in seno all'ITEM, dando visibilità e facilitando l'interfaccia fra le diverse componenti del Laboratorio. Le attività dell'Équipe perseguono due principali obiettivi.

2.1 La comunità scientifica

Si tratta in questa ricerca di esaminare le condizioni e le modalità di uso di Internet nelle discipline umanistiche, di valutare i cambiamenti sul piano epistemologico dell'introduzione delle nuove tecnologie, e in prospettiva di riflettere sulla possibilità di rendere il Web un medium efficace per la comunicazione e le pubblicazioni scientifiche. L'Équipe si è

3

impegnata anche a creare dei modelli d'infrastrutture di ricerca nell'ambito di progetti europei e di un programma nazionale:

GDREplus: "Hyper-Learning" (2004-2006)

COST A32: "Open Scholarly Communities on the Web" (2006-2010)

ScholarWeb (2008-2009)

2.2 Le edizioni scientifiche numeriche

L'ITEM ha avviato la realizzazione di diverse edizioni numeriche: le opere di Nietzsche, i manoscritti di Rousseau e di Renan, la corrispondenza di Proust, i Quaderni di Valery...

Il Pôle Humanités numériques serve come luogo di scambio d'idee, di esperienze e competenze fra le varie componenti e le altre iniziative che si svolgono a livello nazionale e internazionale.

Queste attività si sono inserite nei progetti:

ANR "Optima" (2007-2010)

Il progetto europeo "Discovery" (2006-2009)

Il progetto europeo "Agora" (2011-2014)

Il progetto *eMan* per la pubblicazione elettronica (e genetica) dei manoscritti moderni:

http://eman.hypotheses.org

L'ITEM, rappresentato da Paolo d'Iorio partecipa al Comitato organizzatore del Consorzio CAHIER, avviato nel 2010 dal Centre d'Études Supérieures de la Renaissance dell'Università di Tours: l'iniziativa mira a facilitare lo scambio dei dati e dei metodi nell'ambito dell'edizione scientifica numerica.

e-Man

Piattaforma di edizione elettronica e genetica dei manoscritti moderni

Il 1 aprile 2016 si è svolto un convegno sul Progetto di edizione numerica della Corrispondenza (*Groupe de travail Correspondance, consortium CAHIER*), con la messa a

4

confronto di riflessioni teoriche e di esperienze. In quella occasione sono state affrontate diverse problematiche, tra le quali:

- . come situare l'edizione della corrispondenza rispetto all'autore, all'opera, per quale uso?
- . come descrivere fisicamente l'oggetto? Rispettandone la forma, il contenuto?
- . quale apparato critico?
- . quali e quanti indici?
- . trascrizione o meno del contenuto delle lettere?

Sul sito vengono segnalati altri programmi utili per le edizioni numeriche:

. Neatline

http//neatline.org

consente di "narrare storie" con mappe e cronologie.

NeatLine text: utile per marcare i paragrafi o determinare zone del testo

NeatLineSimile: linearità temporale

NeatLineWaypoints: cartografia punto per punto.

. Scripto

http://scripto.org

. Omeka Developer Documentation

http://omeka.readthedocs.org

. Linged Data

http://archeonum.hypotheses.org/517

. Manuscrits francophones

Questa piattaforma mette a disposizione tutti i documenti dell'archivio della francofonia letteraria. La pubblicazione dei documenti è il risultato del lavoro dell'Équipe Manuscrit francophone dell'ITEM.

I due primi corpus messi a disposizione (responsabile del progetto Claire Riffard):

. *Jean-Joseph Rabearivelo* (Madagascar): diari, corrispondenza, opere poetiche, narrative, teatrali, critiche, storiche)

Si tratta della riproduzione digitalizzata dei testi pubblicati sotto forma cartacea.

I documenti messi in linea (pagine manoscritte, dattiloscritte, foto) sono presentati sotto forma di immagini accompagnate da schede informative e descrittive che formano il *dossier génétique*. È possibile una ricerca per parole-chiave.

. Mouloud Feraoun (Algérie)

L'archivio è costituito da quaderni di appunti (brogliacci) e documenti completati da schede informative e descrittive.

. Les Lettres françaises d'Aragon: laboratoire d'une écriture et d'une pensée des arts

Il progetto in corso di realizzazione prevede un'edizione e l'analisi degli articoli pubblicati da Aragon nelle "Lettres françaises" (1953-1972). L'edizione numerica di questa rivista e il trattamento informatico del contenuto degli articoli coinvolge problematiche riguardanti la storia politica, estetica e editoriale della letteratura francese del XXe secolo, dall'inizio degli anni '50 all'inizio degli anni '70. L'edizione numerica presenta una dimensione genetica evidente legata all'analisi dei manoscritti come tappe del processo creativo.

Il corpus dei contributi alla rivista costituisce il laboratorio di idee dello scrittore, la testimonianza di un'opera *in fieri*, il banco di prova di tante opere successive.

. Bibliothèque patrimoniale numérique de l'École des Mines Paris-Tech

Si tratta della digitalizzazione del fondo bibliotecario patrimoniale di una delle Grandes Écoles d'Ingegneria francesi: l'École des Mines de Paris.

https://patrimoine.mines-paristech.fr

Il fondo molto ampio contiene : diari, corsi d'insegnanti universitari, mappe, statistiche.

Le informazioni tecniche sull'uso di Omeka sonon molto utili:

https://patrimoine.mines-paristech.fr/informations techniques

. Trascrizione e digitalizzazione dei taccuini di lavoro di etnologi http://transcrire.huma-num.fr

3. La base dati MUSE (Manuscrits, Usage des supporte et de l'Écriture)

Si tratta di un nuovo strumento di ricerca codicologica e genetica per lo studio e l'analisi dei manoscritti moderni.

Siccome l'approccio genetico richiede un notevole impegno di osservazione e di descrizione - premesse necessarie alla classificazione dei documenti che verranno a costituire il "dossier genetico" dell'opera considerata – occorreva offrire un modello omogeneo di riferimento e un rigoroso metodo scientifico per la trascrizione dei manoscritti e per l'interpretazione dei processi di scrittura. Trattandosi di creazione letteraria, le esperienze compiute hanno ampiamente dimostrato che l'oggetto intellettuale non va separato dall'oggetto materiale: decifrare e ordinare sotto forma sequenziale le fasi redazionali implicano costanti andirivieni tra il *leggibile* (o *illeggibile*) – cioè *quello* che è scritto o almeno tracciato – e i segni materiali – cioè *come* viene scritto, da quale mano, con quale strumento, su quale supporto, con quale sistema di segnatura, segni di richiamo, numerazione ecc....

L'ITEM ha sviluppato dei modelli descrittivi applicabili ai manoscritti moderni man mano più perfezionati, riguardanti in particolare le caratteristiche materiali. L'informatizzazione dei dati raccolti, si è imposta come indispensabile fin dagli anni '80. Un primo prototipo della base dati, presentato nel 1993, consentiva di archiviare i tipi di carta e le filigrane in base ad una campionatura di manoscritti autografi datati. Tuttavia quel modello di archiviazione non poteva soddisfare tutte le esigenze degli specialisti di codicologia e di critica genetica che richiedevano una schedatura esaustiva delle occorrenze.

La fase successiva di quel lavoro d'équipe è sata dedicata a rispondere ad una sfida metodologica complessa e originale: rendere tecnologicamente più efficiente e rigorosa la raccolta dei dati, sottoporre a norme precise e omogenee i modelli di descrizione genetica, perfezionare le forme di edizioni elettroniche.

MUSE è ormai operativa dal 2000. Il settore codicologico sviluppa il metodo sperimentato nella prima base codicologica, *Profil*. I dati raccolti riguardavano le informazioni raccolte per ogni tipo di carta filigranata, la descrizione di essa, la collocazione del supporto e la data di utilizzo, ma anche ogni altra caratteristica della carta. Negli anni '90, si erano viste

altre iniziative analoghe, ma si trattava per lo più di tentativi di indicizzazione con la selezione di determinate occorrenze, oppure dello studio approfondito di un solo corpus letterario o filosofico.

Quel tipo di progetto rimaneva tuttavia limitato, tanto più che già nell'Ottocento, ma ancora di più nel '900, si sarebbe diffuso l'uso di carte non filigranate, d'altra parte si poneva il problema del trattamento delle carte isolate, dei frammenti, dei fogli legati in partenza e poi separati in seguito a diverse manipolazioni, o a seconda delle fasi della genesi. Come riuscire a collegare i dati ottenuti in base all'analisi materiali con le osservazioni complementari riguardanti l'organizzazione della superficie grafica?

Nell'ambito di un progamma del CNRS, *Archives de la création*, un nuovo progetto è stato delineato intorno all'obiettivo di creare una base dati relazionale tale da consentire l'approccio dell'oggetto manoscritto da più angolazioni complementari, anche se l'analisi materiale rimane l'asse principale della ricerca. Non si trattava soltanto di offrire un duttile strumento di indagine, ma di integrare la descrizione materiale (tipo di supporto, strumento scrittorio, inventario delle occorrenze) con la raccolta dei dati riguardanti la tipologia delle scritture, la classificazione e l'utilizzo dei fogli, le caratteristiche delle cancellature e della paginazione, in modo da identificare le successive fasi compositive.

Specificità della Base Dati MUSE

Mentre la maggior parte delle basi dati nelle Scienze umane sono prevalentemente informative (schedatura, indicizzazione, repertori...), agevolate dal ricorso a Internet (Reti di biblioteche...), *MUSE* è principalmente uno strumento di studio che intende costruire il proprio oggetto ossia il manoscritto letterario: quindi, non solo uno strumento di ricerca ma anche e soprattutto un sistema euristico che permette di trattare l'analisi dei dati codicologici e genetici, e di elaborare ipotesi interpretative.

Per una descrizione della concezione tecnica della base dati (scelta del programma, sistema relazionale, struttura della base, relazione tra i files) si veda il file annesso (in francese).

Per la presentazione dettagliata dell'intero progetto e il trattamento dei corpus dei manoscritti di Montesquieu, Stendhal, Duchamp, Condorcet, si rimanda all'articolo completo: http://www.item.ens.fr/index.php?id=223455

L'Équipe Techniques et Pratiques de l'écrit (TPE) ha sviluppato la base dati *MUSE*, giunta ormai al suo livello MUSE 7 beta (serge.linkes@univ-lr.fr)

4. Modalità e programmi di trascrizione per le edizioni elettroniche dei manoscritti.

Trascrivere e codificare simultaneamente

Si può trascrivere un manoscritto utilizzando un programma specifico (*Word*, *Open Office*, ecc.). Dopo di che si passa alla tappa della codifica con un programma di codifica professionale tipo *Oxygen*. La codifica permette di strutturare il documento per dargli una forma simile a quella della pagina del manoscritto. La codifica consente di fare ricerche avanzate nel testo: ad esempio se tutti i nomi propri sono stati codificati come nomi di persone in un dato corpus. Se il testo da trascrivere non pone particolari problemi (poche aggiunte infralineari, poche correzioni, poche note marginali, ecc.), è preferibile effettuare la trascrizione prima della codifica.

Invece, nel caso dei manoscritti presenti nella piattaforma *e-Man* - ossia brogliacci pieni di numerosi refusi, correzioni, aggiunte di parole o di frasi infralineari nella pagina o nei margini, ecc... - la trascrizione è una operazione complicata: infatti i programmi di edizione non consentono di riprodurre questo tipo di fenomeno e di trasferirlo online. Se rimane possibile depennare un testo ci sono problemi nel momento del trasferimento informatico online: per cui si rivela necessaria la codifica in un linguaggio specifico. Trattandosi di una edizione elettronica, la possibilità di trascrivere e codificare simultaneamente accorcia molto i tempi.

La necessità di una codifica in XML-TEI adatta ai corpus di e-Man

Il linguaggio più idoneo per la codifica dei documenti letterari è il XML-TEI, il linguaggio XML sviluppato dal Consortium TEI. Esso è particolarmente adatto ai manoscritti: in effetti, il Consortium TEI ha esordito negli anni '80, con un sistema di marcatura (*tags*) per la codifica dei manuscritti antichi e, da allora, non ha smesso di migliorare il sistema per

adattare il XML-TEI alle nuove richieste. Così dal 2011, si trovano molti *tags* utili per la codifica dei fenomeni individuati dalla critica genetica. Tuttavia, si è anche reso necessario adattare e introdurre modifiche per le nostre specifiche ricerche..

Gli strumenti di codifica esistenti

Oltre il programma *Oxygen*, esistono altri strumenti di codifica:

1- Il programma Transkribus (Université d'Innsbruck, Austria

Questo programma *open source* è in libero accesso dal marzo 2015. Prevede un *toolbar* (barra degli strumenti) per la codifica adattabile e un sistema di riconoscimento dei caratteri. È particolarmente efficiente per il corpus di un autore unico. Siccome il nostro corpus contiene un numero svariato di grafie non abbiamo scelto quel programma.

2- Il plug-in Scripto del programma Omeka

Poiché abbiamo costruito la nostra piattaforma con un programma *Omeka*, ci siamo iinteressati al *plug-in* di trascrizione collaborativa *Scripto* proposto da *Omeka*. Esso funziona a partire da un sistema *MediaWiki*. Le trascrizioni compaiono in un *Wiki* di trascrizione. *Omeka* fa comparire le trascrizioni di *Wiki* nella piattaforma. Questo *plug-in Scripto* pone tuttavia un problema di doppia gestione della piattaforma (*Omeka* et *Wiki*) e propone un *toolbar* che non riconosce XML-TEI. Quindi, per ora, abbiamo eliminato tale sistema.

3- Il Wiki di trascrizione del progetto Bentham (University College London)

Il *Transcribe Bentham Project* è un progetto di trascrizione collaborativa dei manoscritti di Jeremy Bentham. È stato elaborato dall'University College London nel 2010 a partire da un *Wiki* di trascrizione del sistema *MediaWiki*. Esso propone un *toolbar* per la codifica in XML-TEI, reso compatibile con i caratteri specifici dei manoscritti di Jeremy Bentham. Questo *Wiki* di trascrizione di UCL è scaricabile gratuitamente ed è *open source*. Se riuscissimo ad adattare il *toolbar* di questo *Wiki* di trascrizione al nostro progetto, potremmo prevederne l'utilizzo in un *Wiki* o direttamente in *Omeka*.

4- Il toolbar Tiny MCE integrato a Omeka

Omeka possiede un *toolbar* in HTML. Si può completare il *toolbar* esistente aggiungendovi nouvi tasti. Siamo riusciti ad aggiungere un tasto « depennare » e una tavolozza di colori. Tuttavia, sembra che il numero di tasti integrabili sia limitato. In definitiva, la scelta dello strumento da utilizzare per creare il nostro *toolbar* di codifica non è definitavemente fissata.

La redazione di un manuale di codifica

Anche se il *toolbar* che vogliamo creare rimane ancora, per ora, al livello di progetto, il lavoro di riflessione che abbiamo svolto ha condotto alla redazione di un manuel di codifica. In quel manuale abbiamo stabilito la lista completa dei *tags* XML-TEI che vorremmo integrare in un *toolbar* di codifica, in particolare dei *tags* idonei per la codifica delle fasi di scrittura caratteristiche nell'analisi genetica.