

*Considerazioni sull'analisi testuale informatica*

Séminaire:  
*Littérature et informatique*

Chiara De Luca  
Università degli Studi di Bologna

## ***Sommario***

Nota introduttiva

1. L'informatica: ausilio o minaccia dell'umanesimo ?
2. L'analisi testuale informatica

## **Nota introduttiva**

Il presente lavoro si propone di introdurre qualche considerazione di carattere generale rispetto alle applicazioni dell'informatica in area umanistica, a partire da uno sguardo sui grandi mutamenti che hanno coinvolto ogni aspetto della comunicazione in un'epoca in cui la trasmissione digitale va prendendo sempre più piede nella vita di un numero crescente di persone.

Nella seconda parte del presente lavoro prenderò in considerazione più da vicino *Wordsmith*, uno dei numerosi programmi di analisi testuale informatica presenti on-line, soffermandomi su alcune delle sue principali possibili applicazioni.

### **1. L'informatica: ausilio o minaccia dell'umanesimo ?**

Capita sempre più spesso di sentire o di leggere interrogazioni più o meno accurate, a volte catastrofiche, concernenti il futuro dell'umanesimo e del libro cartaceo nell'epoca di internet e della comunicazione digitale. Io stessa mi chiedo dove potrà portarci, di qui a pochi anni, il progresso tecnologico che così velocemente ha trasformato il nostro modo di comunicare, di trasmettere e ricevere informazioni, accumulare e consultare dati, ma anche il nostro modo di rapportarci agli altri e di gestire i rapporti interpersonali, se è vero che «[...] nella nostra civilizzazione, nata con la razionalizzazione scientifica moderna, la scelta della tecnica costituisce comunque un elemento di *continuità nel progetto* – progetto in cui l'informatica non è una rivoluzione, ma piuttosto un'intensificazione. Ciò che subisce trasformazione e rottura sono piuttosto i modi dell'apprendimento e della trasmissione del sapere».

La parola scritta ha subito un vero e proprio stravolgimento ad opera di e-mail, SMS e cellulari. La comunicazione è sempre più veloce, ci consente di trasmettere informazioni in tempo reale, quindi le maglie del filtro della scrittura si allargano sempre più, in modo direttamente proporzionale alla velocità della trasmissione via etere e nella rete.

In rapporto al papiro cartaceo, la mail richiede molta meno riflessione, meno tempo. Scrivere al computer facilita le correzioni, le cancellature, le riscritture, le ripetizioni. E alla fine della stesura, mediamente più veloce e meno accorta di quella di

una lettera cartacea, basta ciccicare, e la lettera viene immediatamente recapitata al destinatario. Una volta si sarebbe dovuto scrivere, piegare il foglio, imbustare, sigillare, e cercare la cassetta della posta più vicina. Tutto ciò lasciava un margine maggiore di riflessione, ripensamento, riorganizzazione e rielaborazione dell'informazione. La lingua utilizzata per scrivere una mail è inoltre mediamente più trascurata, meno elaborata e riflettuta.

Lo stesso dicasi per il telefono. La modalità della conversazione via cavo è cambiata moltissimo con l'avvento del cellulare. Mentre prima si parlava al telefono in casa, al massimo in una cabina, forse cercando il momento e la situazione opportuna, adesso si parla al telefono ovunque, in ogni situazione, incuranti delle orecchie e degli occhi altrui. Si parla al telefono guidando, facendo la spesa, camminando, a volte anche quando si è in compagnia di amici. E anche in questo caso, come in quello della comunicazione via e-mail, l'informazione è meno filtrata, più immediata, meno riflettuta. Spesso non abbiamo il tempo di rielaborare noi stessi quello che ci sta succedendo e che stiamo raccontando, perché la narrazione avviene in tempo reale. Il nostro interlocutore sta vivendo con noi la situazione in cui ci troviamo, nello stesso momento in cui la viviamo noi. Il fatto che il cellulare ci raggiunga ovunque rende inoltre spesso necessario l'utilizzo di una sorta di codice, di un linguaggio allusivo o ambiguo, se si desidera che gli altri non sappiano di cosa esattamente si stia parlando. In altri casi, la fretta costringe ad una comunicazione veloce, concitata, essenziale.

E la comunicazione si riduce alla sua massima essenzialità nel caso dell'SMS. Avendo pochi caratteri – e, spesso, poco tempo – a disposizione, si tende ad accorciare le parole, utilizzando abbreviazioni, simboli, elisioni, che stanno diffondendosi e uniformandosi, al punto che va via via creandosi una specie di codice cifrato, il cui fine principale è quello di velocizzare la conversazione.

In breve, tutte le nuove forme di comunicazione tendono a semplificare e velocizzare la trasmissione del messaggio, e a uniformare la lingua di trasmissione. Utilizziamo sempre di meno la carta, curiamo sempre meno la qualità del messaggio, a vantaggio della sua funzione, e tendiamo ad economizzare sui tempi destinati alla riflessione. A volte si ha l'impressione che il messaggio stesso, al pari della lingua, subisca una sorta di omologazione e semplificazione, che va a discapito dell'originalità. Si potrebbe dire che la lingua in molte occasioni si spersonalizza, che il messaggio si fa

più freddo ed essenziale, quasi adattandosi al suo mezzo di divulgazione. Ma non è sempre così. Il tutto, a mio parere, sta nel tenere a mente il nucleo originario della comunicazione, che è il messaggio stesso. Se si privilegia il contenuto in rapporto alla forma, la forma diviene effettivamente veicolare, e la velocità e immediatezza della trasmissione del messaggio diviene ad esso funzionale.

Per quanto, spesso, mi succeda di avvertire un impoverimento della comunicazione nel mio quotidiano, non penso che la comunicazione digitale potrà mai sconvolgere o abbattere interamente le modalità tradizionali di trasmissione dell'informazione.

Allo stesso modo, non credo che l'ipertesto potrà mai sostituire il libro cartaceo, e che l'umanesimo sarà mai posto in un ruolo subordinato nell'epoca dell'informatica. Credo che le connotazioni affettive - a volte anche vagamente sentimentali - del libro cartaceo non potranno mai essere sostituite dallo schermo illuminato di un computer. È vero che il superamento della linearità del testo può favorire l'interazione tra scrittore lettore, che i ruoli tradizionali possono in qualche caso essere parzialmente ridefiniti, ma il testo elettronico prende le mosse dal libro tradizionale, che ne è presupposto e sostegno. Ma, considerando più da vicino l'ipertesto, si può dire che esso in fondo si basa su un principio non dissimile da quello del libro tradizionale «la struttura ipertestuale infatti dà l'illusione della creatività, della scoperta di legami tra fenomeni diversi, che costituisce il primo passo di ogni ricerca innovativa: ma si tratta, appunto, di una illusione. L'utilizzatore trova "da sé", cioè, crede di trovare, solo i legami ipertestuali che i compilatori hanno già previsto, senza che si possa aggiungere nulla. La ricerca scientifica è invece altra cosa, è soprattutto proprio il percorrere strade nuove, fuori dai sentieri battuti». La creatività, la ricerca della novità, l'individuazione di piste non ancora battute dalle ricerche precedenti continuano a restare prerogativa dell'uomo, ed è su queste modalità che si organizza il progresso informatico, non viceversa. La macchina non inventa nulla, difficilmente scopre qualcosa di nuovo a prescindere dall'analisi critica dei dati compiuta dall'uomo, e ogni eventuale novità necessita comunque di verifica da parte del ricercatore.

Differente è il caso del libro in CD Rom, che si sta diffondendo sempre di più, e che può rappresentare davvero uno strumento utile all'avanzamento della ricerca. Questo perché il CD Rom consente di immagazzinare una quantità enorme di testi, che

sarebbe altrimenti difficile poter avere a disposizione tutti contemporaneamente. Esso agevola inoltre l'individuazione di citazioni, punti specifici, il chiarimento della provenienza di eventuali reminescenze, la cui contestualizzazione e verifica richiederebbe molto tempo e lavoro, in vista di un risultato non sempre garantito. Ritengo che, in questo caso, cartaceo e multimediale possano rafforzarsi a vicenda, ma che l'uno continui ad essere punto di partenza e di arrivo dell'altro.

Lo stesso dicasi per il rapporto tra archivio tradizionale e archivio multimediale. Il vantaggio del secondo rispetto al primo sta nella possibilità di immagazzinare una quantità potenzialmente infinita di informazioni, la cui gestione si è fatta più semplice e più veloce. Anche in questo caso, le strutture gestionali non sono pensabili prescindendo dal principio di base: l'esigenza di una suddivisione e organizzazione dei saperi, principio radicato molto prima nel tempo rispetto all'inizio del boom informatico.

Anche per quanto riguarda l'analisi testuale informatica, ritengo che la strada da seguire sia quella di una integrazione delle metodiche. Non credo che l'analisi informatica potrà mai sostituire l'uomo, ma credo che possa rivelarsi molto utile nel caso in cui ci si trovi a gestire una grande quantità di informazioni, un *corpus* esteso di testi. In tal caso, il computer potrebbe costituire un ausilio delle capacità umane, economizzando sugli spazi e i tempi dedicati alla parte «compilativa» della ricerca, a vantaggio della parte innovativa. Si tratterebbe comunque di un ausilio di tipo compilativi, poiché siamo ancora bel lontani dalla simulazione dell'intelligenza umana. Anche gli esperimenti nell'ambito della «intelligenza artificiale» compiuti fino ad ora, non hanno infatti ancora fornito risultati soddisfacenti al punto da far sperare – o temere – che il personal computer possa sostituirsi al suo fruitore, spersonalizzandolo. Ciò non toglie che lo strumento informatico possa rivelarsi anche molto utile nello svolgimento di operazioni che l'essere umano potrebbe sì compiere, ma con un dispendio di tempo e di energie molto maggiori, da sottrarsi alla ricerca. Il computer potrebbe quindi rivelarsi uno strumento utile all'avanzamento del sapere nel senso di una velocizzazione e uno snellimento nella registrazione dei dati.

L'utilizzo della ricerca informatica dovrebbe a mio parere essere funzionalizzato all'attività umana, non diventarne un surrogato, e, perciò, un ostacolo. Le possibilità che offre oggi l'informatica sono molteplici, e, se sfruttate in maniera intelligente e

consapevole, possono agevolare notevolmente l'avanzamento della ricerca, costituendo un supporto per l'attività umana. Internet, per esempio, rappresenta una banca dati immensa, potenzialmente infinita, che riduce i tempi e gli spazi della gestione e amministrazione dei dati. Attraverso i motori di ricerca si possono ricavare molto velocemente informazioni il cui reperimento, fino a pochi anni fa, avrebbe richiesto grande dispendio di tempo ed energia, all'occorrenza anche spostamenti e lunghe ricerche sul luogo. Ma i risultati di una ricerca informatica debbono essere sempre accuratamente vagliati e distinti in base all'importanza, all'attendibilità. Ed è qui che entra in gioco l'attività umana, che non può ancora essere sostituita. Sta al ricercatore distinguere tra i dati che possono essere utili alla sua ricerca e quelli che, invece, si rivelano superflui, o addirittura fuorvianti.

Il vantaggio di internet risiede quindi principalmente nella velocità delle informazioni e nel coordinamento delle stesse. Il pericolo è quello di sopravvalutare il peso scientifico di alcune informazioni, affidandosi ad esse ciecamente, peccando di spirito critico.

Credo che il termine «intelligenza artificiale» metta già di per se stesso in guardia rispetto ai rischi legati all'idea che l'analisi informatica possa sostituire le capacità umane. L'intelligenza di cui si parla è «artificiale», essa presuppone cioè l'intelligenza umana, di cui rappresenta una sorta di emulazione, di necessità imperfetta che, tuttavia, se sfruttata al massimo delle sue potenzialità, nella costante consapevolezza dei suoi limiti, può rappresentare un valido supporto all'attività umana. Rispetto alla intelligenza artificiale e alle sue funzioni e possibili utilizzazioni, esistono – al di là delle molteplici sfumature possibili – due linee di pensiero predominanti:

In una prima accezione l'intelligenza artificiale sarebbe quell'area delle scienze computazionali che intende esplicitamente scoprire e analizzare a fondo i meccanismi del pensiero, della creatività e del linguaggio, di quegli elementi, cioè che fanno degli uomini degli esseri intelligenti. In una seconda accezione, l'intelligenza artificiale sarebbe quel settore delle scienze computazionali specificamente finalizzato a costruire sistemi intelligenti ovvero che esplichino le proprie attività in un modo che potremmo definire «intelligente». Tali definizioni caratterizzano di fatto due approcci tra loro ben differenti, ma coesistenti all'interno di un unico seppur complesso e articolato ambito di ricerca.

Queste due linee di pensiero corrono parallele, giungendo ad intersecarsi in alcuni punti, a scontrarsi in altri. Credo che per entrambe esista una valida giustificazione, e che della loro integrazione e del reciproco rafforzamento possa nascere l'attitudine più utile e meno rischiosa nei confronti della simulazione dell'intelligenza umana. L'obiettivo potrebbe essere uno strumento che simuli, per quanto possibile e nella misura in cui sono a noi noti, i meccanismi mediante i quali la mente umana opera, visto come ausilio, più che protesi o sostituto, dell'intelligenza umana.

Tale strumento può essere così utilizzato per compiere in modo «economico» e «intelligente» tutte quelle azioni che richiederebbero un grande dispendio di tempo e di energie che verrebbero sottratti allo svolgimento di tutte quelle attività che soltanto la mente umana è in grado di compiere:

Il problema della generazione ovvero della produzione di testi coerenti, così come quello della ricezione e della comprensione reale del linguaggio naturale evidenzia infatti la precisa esigenza di un sorprendente investimento di tempi e di energie da destinarsi ad un tipo di ricerca assai complesso, di natura sicuramente interdisciplinare e trasversale ad un tempo stesso, delineato al fine di giungere a risultati anche minimamente soddisfacenti.

È in quest'ottica che si muove la cosiddetta «ipotesi debole», che ha come obiettivo la parziale simulazione di attività cosiddette «intelligenti», e chi si contrappone la cosiddetta «ipotesi forte», che si ripropone invece l'ambizioso obiettivo di riprodurre il funzionamento delle attività del cervello umano. In tale ottica, «il ruolo dell'umanista esperto anche di modellistica computazionale diviene fondamentale al fine di progettare prassi di ricerca che possano rispecchiare di fatto la complessità del pensiero umano e dell'individuo in tutte le molteplici componenti e caratteristiche che lo contraddistinguono». Un atteggiamento di chiusura che neghi l'importanza che lo strumento informatico può assumere anche per quanto riguarda una ricerca in campo umanistico, risulterebbe quindi sterile. Uno dei problemi derivati dalla suddivisione – se non parcellizzazione – dei saperi risiede spesso nella mancata cooperazione dell'esperto informatico e dell'umanista, che determina il prevalere delle funzioni dell'uno rispetto a quelle dell'altro, in luogo di una più produttiva reciproca integrazione delle stesse. L'ideale sarebbe la condivisione delle informazioni e la reciproca apertura, e la

creazione di «figure miste», che siano in grado di superare, o almeno ridurre la distanza tra i saperi, perché, più che di una prevalenza dell'uno rispetto all'altro, si possa parlare di una reciproca integrazione, che determinerebbe, di necessità, un rafforzamento in entrambe le direzioni. Questo ridurrebbe il rischio di una spersonalizzazione dell'individuo, conseguenza possibile di una sua funzionalizzazione alla macchina, che dovrebbe esserne invece nient'altro che strumento, mezzo, quindi, di arricchimento e valorizzazione, più che di impoverimento e svalutazione delle facoltà umane:

Nella nostra intrinseca prospettiva, occorre soprattutto curare – e non sembri un gioco di parole – che il *personal* non spersonalizzi la persona; che lo strumento, con uno spostamento della categoria mediale a quella finale, non venga a sussumere sulla propria funzione quel *per sé* che abbiamo riconosciuto appunto tra le cifre essenziali dell'uomo umanistico; che la macchina, prodigiosamente capace di lavorare sulla *quantità*, non tenda a svuotare delle sue qualità chi la utilizza; che la «persona, facendosi tutt'uno con la sua protesi informatica, non finisca per restare impigliata nella Rete (con la R maiuscola), dispersa o addirittura disintegrata nella globalizzazione mediatica».

Ritengo che, per quanto riguarda l'analisi profonda del testo letterario, siamo ancora ben lontani dalla possibilità di svolgerla interamente a mezzo di un programma informatico, per quanto sofisticato esso possa essere. L'analisi informatica può senz'altro rappresentare un valido strumento di ausilio, ma la parte più importante, quella della verifica e della lettura dei dati, spetta al ricercatore, che la compirà sulla base della propria esperienza e delle proprie acquisizioni, poiché «in una prospettiva di ricerca cognitiva e computazionale, un testo linguistico costituisce una entità avente struttura complessa, sintattica, semantica e pragmatica precisamente, la cui percezione, che si realizza appunto attraverso i processi di lettura, scatena, da parte del lettore umano o artificiale, nel caso si tratti di un sistema di simulazione, tutta una serie di processi di selezione, organizzazione e sintesi delle informazioni contenute nel testo, processi basati sulle rappresentazioni che il lettore possiede e sulle capacità da parte del lettore stesso di attivazione dei medesimi, secondo i criteri della pertinenza e della economicità».

## 2. L'analisi testuale informatica

L'analisi testuale informatica può senz'altro essere un ausilio molto utile al lavoro del ricercatore in materie umanistiche. Ma, appunto, è un ausilio, un supporto che aiuta a svolgere più velocemente un lavoro che potrebbe essere svolto anche manualmente, ma con un grande dispendio di energia e un alto margine di errore. I dati risultanti dall'analisi testuale debbono comunque essere sottoposti a verifica da parte del ricercatore, sulla base delle proprie competenze ed esperienze. L'analisi informatica ha inoltre il vantaggio di favorire la gestione di un'enorme quantità di testi, in cui il ricercatore rischierebbe di smarrirsi e di facilitare la individuazione di parole, ricorrenze, percentuali, che richiederebbe altrimenti molto tempo, da sottrarsi all'analisi e alla critica dei risultati statistici, attività che, allo stato attuale della ricerca informatica, sono ancora territorio privilegiato dell'uomo, di cui la macchina non è ancora in grado di simulare le funzioni intellettive e analitiche più complesse.

Credo perciò che il compito più importante nell'analisi testuale spetti ancora al ricercatore, che, comunque, trae notevoli vantaggi dall'utilizzo razionale della macchina, nella piena consapevolezza dei suoi limiti, e della propria responsabilità di integrarli mediante le proprie facoltà intellettive e la propria intelligenza. È perciò auspicabile che il ricercatore sia aperto e ricettivo nei confronti delle possibilità offerte dall'analisi informatica, per poterle sfruttare al meglio, senza aspettarsi risultati infallibili e definitivi già di per se stessi, bensì una serie di dati su cui è indispensabile lavorare ancora a lungo per trarne delle considerazioni critiche.

Oggi esistono molti programmi di analisi testuale, molti dei quali sono disponibili anche on-line. Essi sono molto utili per l'identificazione delle concordanze, la registrazione delle ricorrenze delle parole, la stesura di elenchi tematici dettagliati.

Il programma *Wordsmith*, per esempio, offre all'analisi testuale diversi strumenti utili e facili da utilizzare. Attraverso una finestra di dialogo, è possibile selezionare uno o più testi da sottoporre ad analisi. Molti testi sono oggi disponibili on-line, ma è anche possibile scannerizzare testi altrimenti non disponibili.

Attraverso il *Viewer Tool* è possibile esaminare i files in vari formati, ma esso può anche essere utilizzato per realizzare una copia del file di testo provvista di numerazione di frasi o paragrafi, o per allineare due versioni di un testo, mostrando

alternativamente i paragrafi e le frasi che li costituiscono.

È inoltre possibile eseguire delle liste di parole singole o di stringhe di parole, fino a otto parole ciascuna. Il procedimento può essere utile per isolare la ricorrenza di determinate parole, o determinate stringhe di parole – che può avere implicazioni anche sul piano tematico. L'analisi delle stringhe di parole può essere utile anche per identificare una concordanza. Identificare la concordanza significa vedere quali sono le parole che ricorrono a destra o a sinistra (o in entrambe le “direzioni”) di una parola data.

All'interno di *Wordsmith*, il programma *concorde* offre diversi strumenti per identificare, allineare e organizzare le concordanze.

Il programma *concorde* si può applicare all'analisi di file di testo DOS, Text only, ASCII o ANSI.

Nell'uso di *concorde* è necessario specificare una parola da ricercare, di cui *concorde* registrerà la ricorrenza nel file o nei file che sono stati scelti.

In tal modo, è possibile registrare non soltanto la ricorrenza di una singola parola, bensì anche il contesto o i contesti specifici in cui ricorre. Il vantaggio è che, avendo davanti tutti i casi in cui una determinata parola è utilizzata, è possibile anche arrivare a chiarirsi meglio la sua funzione e l'eventuale valore simbolico che l'autore le attribuisce. Per quanto riguarda la traduzione, il contesto pragmatico della parola può aiutare a individuarne con maggiore chiarezza il significato nella propria lingua, evitando così fraintendimenti.

Chi invece stia studiando una lingua straniera può servirsi del programma di analisi delle concordanze per apprendere o fissare più chiaramente il contesto d'uso di una parola.

Il passo successivo consiste nella identificazione dei contesti in cui la parola ricorre più frequentemente, è cioè più frequentemente attestata, o giustificata dall'uso.

È anche possibile specificare una parola – o una stringa di parole - che si desidera escludere dal calcolo delle concordanze. Alla parola o alla stringa di parole è poi possibile attribuire un simbolo in modo che non sia necessario ridigitare ogni volta la parola o l'intera sequenza, bensì sia sufficiente ciccare sul simbolo che le rappresenta.

Mediante la funzione *Collocation display* le collocazioni (parole che ricorrono in unione ad altre parole) vengono registrate e mostrate in ordine di frequenza. Oltre a

visualizzare ogni parola, è possibile anche calcolare il totale delle ricorrenze della parola scelta per la ricerca della concordanza e il totale delle ricorrenze delle parole (fino a cinque) a destra e a sinistra della parola scelta. La posizione centrale, quella, cioè, occupata dalla parola scelta, è indicata di volta in volta con un asterisco.

I numeri di riferimento sono:

1. il totale delle ricorrenze della parola trovata nelle vicinanze di quella oggetto di ricerca
2. il totale delle sue ricorrenze alla sinistra della parola oggetto di ricerca
3. il totale delle sue ricorrenze alla destra della parola scelta
4. una serie di frequenze individuali alla sinistra della parola oggetto di ricerca
5. una colonna di asterischi che rappresenta la parola oggetto di ricerca
6. una serie di frequenze individuali alla destra della parola oggetto di ricerca

Il numero delle colonne in 2. e 4. dipende dall'orizzonte di collocamento. L'orizzonte di collocamento consiste nel numero delle collocazioni che concorde troverà alla destra e alla sinistra della parola oggetto di ricerca e la distanza utilizzata da *KeyWords* nella ricerca dei *plot-links*.

In tal modo è possibile calcolare esattamente la ricorrenza di una parola vicino a quella che è oggetto di ricerca e la sua ricorrenza in generale. La parola più frequentemente ricorrente viene segnalata in rosso.

È poi possibile ripetere la ricerca dal principio, variando l'orizzonte della collocazione.

In tal modo è possibile valutare il contesto lessicale e grammaticale caratteristico del ricorrere di una parola.

La funzione *Mutual information* (MI) mette in relazione una parola con l'altra. Il programma mostra non soltanto la frequenza del due parole (la cui relazione è bidirezionale) bensì anche l'occorrenza di una singola parola.

La funzione *Match list* consente poi di effettuare una cernita nelle liste di parole individuate, per riorganizzarle in una nuova lista di parole o in una concordanza che raggruppi determinate parole che presentino le caratteristiche generali indicate dal ricercatore. Esse verranno poi segnalate dal programma, e si potrà scegliere se eliminare

tutte le parole marcate, o mantenerle, a seconda dello scopo della ricerca.

Mediante la funzione *Consistency Analysis (Simple)* è possibile mettere a confronto oltre 50 liste di parole. Con questa procedura è possibile processare un gran numero di testi, elaborando una lista che mostra la frequenza nei testi presi in analisi. Mediante la funzione *Consistency Analysis (Detailed)* è possibile svolgere le stesse operazioni, ottenendo però risultati più dettagliati.

Queste sono soltanto alcune delle operazioni che è possibile svolgere con uno dei programmi di analisi testuale. Credo che queste funzioni siano utili soprattutto perché consentono di compiere operazioni non tanto complesse, quanto dispendiose in termini di tempo e attenzione. Ma in ogni caso, l'intervento del ricercatore, della sua esperienza del testo o dei testi da analizzare è sempre indispensabile. Il ricercatore deve sapere infatti quali sono le parole da cercare, le ricorrenze e concordanze che è utile focalizzare, e, soprattutto, deve sempre avere ben chiaro il fine della ricerca, l'obiettivo iniziale, da cui poi l'analisi si può scostare, senza comunque mai prescindere da esso. Il ricercatore deve poi vagliare l'attendibilità dei dati, e metterli in rapporto alle osservazioni e agli studi da lui stesso compiuti. I risultati della ricerca informatica possono senz'altro aggiungere qualcosa alle sue acquisizioni, indirizzare la sua ricerca in un'altra direzione sulla base di nuovi particolari emersi, ma è sempre l'uomo a decidere i parametri della nuova ricerca.