

## *Indice*

1. Dalla scrittura alla testualità .....	1
2. I livelli del testo .....	7
3. La codifica TEI ( <i>Text Encoding Initiative</i> ) .....	15
4. La preparazione del testo elettronico .....	31
Riferimenti bibliografici .....	37
Sitografia .....	41

Oggi, chi si avvicina al computer per la prima volta ha sempre più la sensazione di trovarsi di fronte ad una macchina che serve per *scrivere testi*. La concezione tradizionale del computer come strumento che serve per *fare calcoli* ha via via lasciato il posto ad una nuova prospettiva, considerata ormai la grandissima diffusione della cosiddetta 'scrittura elettronica' che, nelle scuole e negli uffici come nelle case, consente di produrre e trasmettere documenti di tutti i generi o di alimentare ed interrogare le banche dati che usiamo quotidianamente, per compiere operazioni bancarie o conoscere l'orario dei treni.

È facile constatare, però, che l'espressione *scrivere testi* nasconde una fondamentale ambiguità: infatti, la *scrittura* elettronica è un'operazione diversa dall'elaborazione di un *testo* elettronico, e questo, in una società che richiede conoscenze sempre più approfondite e precise, deve essere ben chiaro a chi si accosta al computer. In questo lavoro illustreremo in modo molto sintetico cosa significhi costruire un *testo elettronico*, quale impegno richieda e, in ultima analisi, perché tutto ciò sia diventato sempre più importante nella nostra *era informatica*.

## 1. Dalla scrittura alla testualità

Chi ha avuto modo di usare un programma come *Word* di Microsoft ha già chiare le caratteristiche essenziali dei sistemi moderni di scrittura elettronica. Queste applicazioni consentono di aprire un qualunque archivio (*file*) e di inserirvi una qualunque sequenza di caratteri, che sarà poi trasformata dal computer in una sequenza di *bit* (0 e 1), invisibili all'utente ma presenti nella memoria elettronica. Il calcolatore è una macchina che può riconoscere simboli (ad esempio, caratteri alfabetici e numerici) e operare su di essi, per registrarli, riprodurli o compiere elaborazioni. L'utente introduce dati attraverso dispositivi di *input* (una tastiera, un *mouse* o un microfono) e riceve dati attraverso dispositivi di *output* (un video, una stampante o un altoparlante). Il processo che avviene all'interno del computer tra *input* e *output* (che, tecnicamente, si può definire di 'traduzione') non viene colto dall'utente e sembra non riguardarlo più di tanto.

È sufficiente un piccolo esempio per farci capire l'estrema varietà di testi che possiamo scrivere in un *file*, senza contare la possibilità di variare caratteri, colori, stili e così via. Nella fig. 1 alla pagina successiva sono rappresentati schematicamente tre *file* con il loro contenuto.

Il primo *file* (che chiameremo **uno**) contiene l'inizio della *Divina Commedia* di Dante, il secondo (**due**) una ricetta di cucina, e, infine, il terzo (**tre**) un esercizio di matematica.

Per il computer i caratteri presenti nei nostri testi sono *tradotti* attraverso codici che li rendono significativi per la macchina, ad esempio il noto codice ASCII (*American Standard Code for Information Interchange*), che assegnerà ad A la sequenza 10000001 e così via.

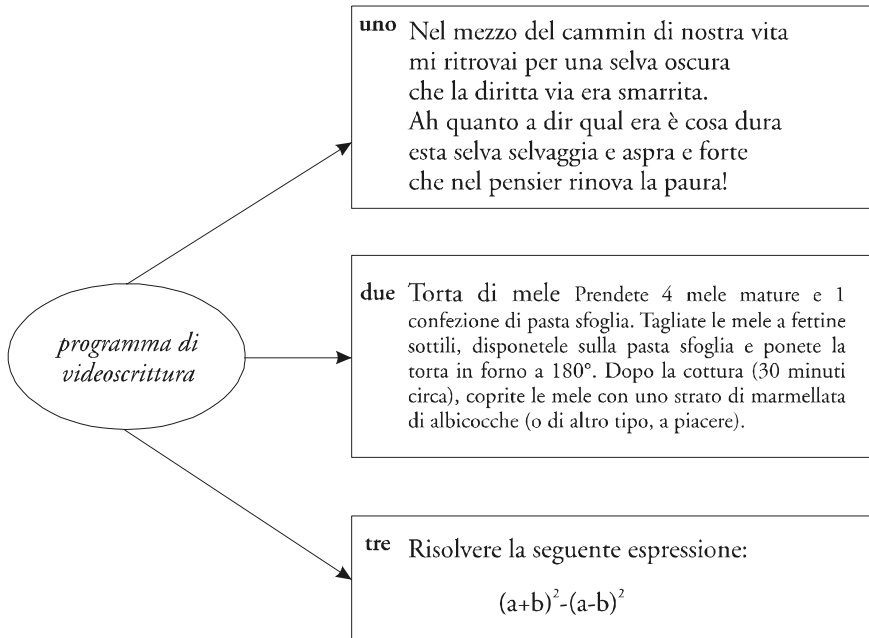


Figura 1.

La situazione non cambierebbe granché per il mezzo elettronico se avessimo in **uno** questa sequenza:

(1)

nelmezzodelcammin dinostravitamiritrovaiperunaselvaoscurachel  
adirittaviaerasmarritaahquantoadirqualeraecosaduraestaselva  
selvaggiaeaspraefortechenelpensierriinoyalapaura

Alcuni caratteri non sarebbero più presenti (lo spazio bianco, le maiuscole ecc.), ma il sistema di codifica e la sua logica rimarrebbero identici. In ogni caso siamo di fronte ad un testo che tecnicamente viene definito

‘leggibile dalla macchina’ (in inglese, *Machine Readable Form*, abbreviato MRF)<sup>1</sup>.

Per l’utente umano le cose vanno, però, molto diversamente. Poiché l’utente di un calcolatore è anche (e soprattutto) un parlante, che conosce una lingua naturale (o più lingue) ed è in grado di comprendere e produrre *testi* orali e scritti, la differenza tra le due *sequenze* registrate in **uno** è estremamente significativa. Come hanno mostrato sempre di più le ricerche di linguistica e semiotica, il parlante non produce semplici concatenazioni di suoni o caratteri, ma unità strutturate, espressioni che hanno una coesione formale e una coerenza concettuale.

Queste informazioni sono del tutto assenti nella logica della macchina. Così, l’*a capo* nel testo dantesco identifica per il lettore (umano) la fine di un verso o la maiuscola l’inizio di un nuovo periodo, mentre per il calcolatore si tratta di semplici codici diversi tra i tanti disponibili. In più, naturalmente, soltanto il lettore è in grado di porsi il problema dell’*autore* del testo come diverso dall’*autore* del *file*: il computer registrerà soltanto l’utente che ha ‘aperto’ un determinato archivio elettronico.

Se consideriamo ancora un altro esempio, ci accorgiamo di quanto sia radicale la differenza tra sequenze e testi. Prendiamo nuovamente una sequenza in cui siano state omesse praticamente tutte le convenzioni dell’italiano scritto:

(2)

carneade1carneadedicireneiisecacfilosofoeoratorechieracostuairu  
minavatrasedonabbondiosedutosulsuoseggioloneinunastanzadelpi  
anosuperioreconunlibriccioloaperto davantiquandoperpetuaentro  
aportarglilimbasciata2ditoniocheconunpretestoavevachiestodipar  
lareadonabbondio

e confrontiamola con il testo originale:

<sup>1</sup> La sigla MRF è usata normalmente nella letteratura tecnica e scientifica e sarà usata anche in questo volume.

(3)

- Carneade!<sup>1</sup> Chi era costui? - ruminava tra sé don Abbondio seduto sul suo seggiolone, in una stanza del piano superiore, con un libricciolo aperto davanti, quando Perpetua entrò a portargli l'imbasciata.<sup>2</sup>

---

1. Carneade di Cirene (II sec. a.C.), filosofo e oratore. - 2. di Tonio, che con un pretesto aveva chiesto di parlare a Don Abbondio.

L'esempio è tratto dal volume XIV de *Le pagine della letteratura italiana*, raccolta di Giuseppe Lipparini (Milano, Signorelli editore, 1951, pag. 151), e riporta un brano assai noto del capitolo VIII dei *Promessi sposi* di Alessandro Manzoni. Nell'originale, il lettore sa distinguere agevolmente la presenza di un testo complesso, in cui sono presenti le parole di due autori: sopra la linea continua c'è il testo scritto da Manzoni, mentre al disotto quello scritto da Lipparini. Essi sono collegati da un legame che diremmo oggi 'ipertestuale': i due esponenti delle note permettono, infatti, di passare da un testo all'altro in modo diretto.

Il lettore è in grado di riconoscere la coerenza del testo, cogliendo la presenza di autori diversi e l'architettura delle varie parti; ad esempio, potrà distinguere tra l'esponente '2' presente nel *file tre* precedente (che indica 'l'elevamento a potenza') da quello in (3) che funge da àncora ipertestuale.

Queste brevi considerazioni ci portano a concludere che, in ogni testo scritto, le cosiddette 'convenzioni' ortografiche e tipografiche rappresentano il risultato di complesse operazioni *metalinguistiche*, cioè di analisi di una lingua naturale, dei suoi usi oltreché del suo sviluppo storico. La delimitazione delle parole e dei periodi, la marcatura dei nomi 'propri' rispetto a quella dei nomi 'comuni' o delle sigle, dei dialoghi rispetto alle esposizioni monologiche – soltanto per citare alcune delle particolarità più evidenti – derivano dall'esercizio della capacità metalinguistica che è presente in ogni parlante.

Di fronte alla sequenza *Franco* registrata in un *file*, in nessun caso il computer sarà in grado di ‘capire’ se ha di fronte un ‘nome proprio’ o un ‘cognome’ o, ancora, un ‘aggettivo’ posto dopo un punto fermo ad inizio di periodo. Tutto ciò è riflesso nella funzionalità delle operazioni *Trova* e *Sostituisci*, che possiede ogni *text-editor*. La ricerca di sequenze da trovare o sostituire è guidata da criteri quali il tipo e la grandezza del carattere, non dal significato o dall’appartenenza a categorie grammaticali.

Ancora, se siamo di fronte a queste due frasi italiane:

(4)

a. *C'erano lunghe file davanti ai negozi.*

b. *Ho aperto un file di testo.*

ci accorgiamo che in (b) c'è un cambio di lingua; la parola *file* è un cosiddetto ‘prestito’ dall’inglese e non ha niente a che vedere con l’italiana *file*, che pure si scrive in modo identico. Per la funzione *Trova e Sostituisci* le due sequenze non sarebbero distinte, ma trattate analogamente.