

PER UNA EDIZIONE INFORMATICA DEI *MOTTETTI*  
DI EUGENIO MONTALE: VARIANTI E ANALISI STATISTICA.

Il problema della variantistica, da sempre trattato nell'ambito filologico-letterario, presenta molteplici e discussi problemi metodologici che oggi è possibile affrontare avvalendosi dell'aiuto delle tecnologie informatiche. Queste offrono infatti allo studioso la possibilità di una verifica immediata di tutti gli elementi analizzati, nonché un lavoro computazionale e di collazione più rapido e soprattutto più preciso. Ma rilevante è il fatto che

nel procedimento ecdotico pre-informatico la trascrizione è un momento 'debole', privo di problemi teorici e, al limite, non scientifico; non a caso per lungo tempo, specie prima della diffusione dei microfilm, esso era affidato ad aiutanti, a corrispondenti locali, a bibliotecari, etc. Non era la trascrizione e neppure la *collatio*, bensì l'*emendatio* il luogo in cui erano chiamati ad esercitarsi l'acume e la dottrina del filologo.

Al contrario, nel procedimento ecdotico informatico la trascrizione è intesa come codifica, rappresenta dunque un momento 'forte', anzi decisivo dell'intera operazione [...] perché da una corretta codifica dipendono tutte le successive possibilità di trattamento informatico del testo.<sup>1</sup>

Da questa impostazione ha preso avvio il nostro studio delle varianti e di analisi statistica, dando al testo preso in esame una codifica appropriata affinché lo stesso potesse essere 'processato'. Per la codifica del testo sono stati utilizzati due linguaggi di marcatura: lo SGML (*Standard Generalized Markup*

---

<sup>1</sup> R. Mordenti, *Oltre il gesto della moglie di Lot: esperienze e proposte di uso dell'informatica per l'edizione critica*, in Claude Cazalé Bérard (éd. par), *Récit et informatique. Actes de la journée d'études C.R.L.L.I.*, Université de Paris X-Nanterre, La Garenne-Colombes, Editions de l'Espace Européen, 1991, pp. 37-52.

*Language*)<sup>2</sup> secondo le regole del progetto TEI (*Text Encoding Initiative*),<sup>3</sup> e il linguaggio COCOA (*Computer Oxford Concordance On Atlas*) per il sistema TACT (*Text Analysis Computing Tools*).<sup>4</sup> Il testo base preso in esame, sia per costruire il documento SGML seguendo le regole del progetto TEI, sia per effettuare lo studio statistico con il sistema TACT, è stato estratto da: Eugenio Montale, *L'Opera in versi*, edizione critica a cura di G. Contini e R. Bettarini, Einaudi, Torino, 1980. Esattamente è stata esaminata la II sezione della raccolta *Le occasioni* dal titolo *Mottetti*. Il testo appena citato è stato utilizzato sia per la memorizzazione nell'elaboratore elettronico, sia come riferimento per lo studio delle varianti.

I *Mottetti* sono venti brevi poesie scritte tra il 1934 e il 1940, a cui se ne aggiunge una, *Il balcone*, del 1933, che nell'ambito dell'intera raccolta viene collocata come poesia d'apertura e all'esterno delle quattro sezioni in cui è divisa la raccolta stessa. La poesia *Il balcone* fa anch'essa parte dei *Mottetti* ed è lo stesso Autore a farcelo sapere in una nota della terza edizione Einaudi, del 1942.<sup>5</sup>

Sono stati considerati 'testimoni' per questo lavoro tutte le diverse edizioni dei *Mottetti* pubblicate singolarmente su riviste e quotidiani a partire dalla data di stesura fino al 1980, le raccolte pubblicate dal 1939 al 1960 e naturalmente la citata *Opera in versi*. Nonostante Contini e Bettarini ritengano definitiva l'edizione della raccolta de *Le occasioni* pubblicata a Milano da Mondadori nel 1960, sono state prese in considerazione anche varianti posteriori a questa data solo in un caso, trattandosi di un'ulteriore raccolta.<sup>6</sup> In tutto sono stati così collazionati 30 testimoni.

<sup>2</sup> International Organization for Standardization, ISO 8879: *Information processing — Text and office systems — Standard Generalized Markup Language (SGML)*, Geneva: ISO, 1986.

<sup>3</sup> Nel 1987 tre grandi associazioni di studi di scienze umane con metodi informatici, la *Association for Computers and the Humanities* (ACH), la *Association for Computational Linguistics* (ACL), e la *Association for Literary and Linguistic Computing* (ALLC), hanno iniziato un progetto di ricerca a livello internazionale cui hanno partecipato studiosi provenienti da tutto il mondo. Il progetto denominato *Text Encoding Initiative* (TEI) intende definire uno standard di codifica unico per tutti, affinché sia possibile lo scambio universale di documenti e testi tra studiosi, enti di ricerca e sistemi informatici diversi.

<sup>4</sup> TACT è un programma, o meglio un sistema di 15 programmi, prodotto dal *Centre for Computing in the Humanities* dell'Università di Toronto per svolgere ricerche testuali ed analisi su opere letterarie.

<sup>5</sup> «*Il Balcone*: fa parte dei *Mottetti*. È stampato in limine per il suo valore di dedica» (E. Montale, *Le occasioni*, Torino, Einaudi, 1942, p. 105).

<sup>6</sup> E. Montale, *Mottetti e altre poesie*, a cura di Margherita Dalmati, Atene, Istituto Italiano di Cultura, 1971. Il testo è stato citato anche da Contini e Bettarini, ed è stato da noi siglato come 'At'.

Ad ogni singolo testimone, per comodità e per necessità di identificazione durante la marcatura e lo spoglio statistico, è stata assegnata una sigla e i mottetti sono stati numerati progressivamente con numeri ordinali, non presenti nelle edizioni sorgente.

### *Costruzione dell'apparato in formato elettronico*

Attraverso l'utilizzo del linguaggio SGML/TEI è stato realizzato un apparato in formato elettronico con le varianti risultanti dalla collazione di tutti i testimoni presi in esame. La memorizzazione dei testimoni è stata effettuata manualmente e ad ogni atto di copiatura è corrisposto un atto di codifica. La decisione di procedere simultaneamente nelle due operazioni è dipesa dalla scelta di una delle modalità di codifica che il progetto TEI prevede per la costruzione dell'apparato elettronico. Tali modalità sono:

— il metodo di 'riferimento posizionale', in cui viene indicato dalla codifica il blocco di testo dove occorrono una o più varianti, senza specificare puntualmente a quali parole ci si riferisca. Con questo metodo l'apparato può essere collocato sia all'interno che all'esterno del testo, ma non è possibile ricostruire automaticamente le lezioni dei vari testimoni. Questo significa che non vengono sfruttate pienamente le potenzialità che il trattamento informatico offre per l'edizione scientifica di un testo;

— il metodo di 'collegamento a due punti limite' in cui, attraverso la codifica, viene segnalato sia l'inizio che la fine del segmento di testo in cui compaiono delle varianti. Anche in questo caso l'apparato può essere collocato sia all'interno che all'esterno del testo;

— il metodo della 'segmentazione parallela' in cui l'apparato è sempre interlineare e va collocato quindi *obbligatoriamente* all'interno del testo.

Quest'ultimo è stato il metodo utilizzato e la scelta è stata dettata dalla possibilità di trascrivere e codificare le varianti (segnalando con la marca apposita la parola, il carattere, il segno diacritico cui ci si riferisce) e nello stesso tempo di avere davanti, nell'atto della codifica, il testo nel suo scorrere. Quindi c'è stata la possibilità di sapere, nel momento della trascrizione e codifica, da quale edizione è stata tratta la variante e quali edizioni contengono parti di testo comuni a tutti i testimoni. Infine il metodo della 'segmentazione parallela' riesce a soddisfare perfettamente l'esigenza di un apparato costruito con testimoni editi a stampa e manoscritti tra i quali siano presenti varianti 'annidate'<sup>7</sup> all'interno

<sup>7</sup> Si intende per 'variante annidata' un carattere, termine, parola o frase, che sia stata, in uno stesso manoscritto, sostituita più volte dall'autore. Il termine è mutuato direttamente dall'informatica in cui si usano procedure annidate, cioè una interna all'altra.

dello stesso manoscritto. Una codifica così effettuata può facilitare la ricostruzione automatica delle varie versioni di un testo.

Questo metodo è auspicabile che sia utilizzato in tutte quelle occasioni in cui ci si trovi a codificare un testo del quale non si voglia però, o non si possa, scegliere un testo base su cui condurre la collazione. Anzi possiamo dire che con questo metodo la funzione del testo base potrebbe scomparire se considerassimo ogni parola o segno variato una semplice alternativa nei confronti di un'altra parola o segno piuttosto che una scelta di sostituzione da parte dell'autore, ovvero se considerassimo le varianti come versioni alternative del testo preso in esame. Questa assoluta indifferenza del testo base riveste, come si comprende, una grande importanza nel caso di un'edizione critica *in fieri*. Ovviamente nel nostro caso la collazione è stata fatta su una base di dati certa per cui per il confronto variantistico abbiamo sempre fatto riferimento all'edizione critica *L'opera in versi*, accettata dallo stesso Montale in vita.

Nel codificare le varianti si è cercato di dare la massima informazione anche riguardo ad elementi che normalmente, in uno studio filologico-variantistico tradizionale, non vengono considerati. Ciò è dovuto anche al tentativo di ricreare un testo nel formato elettronico mantenendo il più possibile tutte le caratteristiche proprie del testo cartaceo di origine. Quindi oltre i fenomeni lessicali e diacritici, si sono codificati come varianti anche alcuni fenomeni specificamente tipografici.

È stata insomma considerata 'variante' nei testimoni collazionati:

- la presenza o assenza o sostituzione, in uno stesso luogo, di segni diacritici;
- la sostituzione o eliminazione o lo spostamento di singole parole, di sintagmi, di intere frasi;
- la modifica ortografica, in uno stesso luogo, di singole parole;
- la differenza tra le versioni di uno o più testimoni di uno stesso mottetto, della modalità grafica-tipografica.<sup>8</sup>

La gerarchia del linguaggio SGML progetto TEI prevede inoltre, attraverso alcune regole strutturali per le varianti in sequenza, anche la codifica di una struttura 'annidata' delle parti di un testo. Ad ogni mottetto è stato dato oltre al numero progressivo una identificazione <id>, necessaria per le possibili referenze incrociate con elementi presenti nello stesso file o in un file esterno. Lo stesso è stato fatto nel codificare le stanze delle varie poesie. Questi *links* possono riferirsi a note, commenti oppure immagini.

---

<sup>8</sup> Nella codifica per il testo elettronico questa differenza è stata, per comodità, sempre denominata 'tipografica' anche nel caso di un manoscritto.

Per rendere più comprensibile ed esauriente il lavoro è stato costruito l'apparato di ogni singolo mottetto sia nella sua veste tradizionale, sia in quella elettronica. Per fare un esempio, nella Tavola 1 è riportata la codifica per la costruzione dell'apparato elettronico di una parte del verso n. 4 del mottetto *Perché tardi? Nel pino lo scoiattolo*:

```

<l n='4'>
  <app type='tipografica, sostanziale'>
    <rdg wit='L2' hand='EM'>
      <app type='sostanziale'>
        <rdg varSeq='1' type='soppressa'><delhand='EM'
type='depenamento'>nel</del></rdg>
        <rdg varSeq='2' type='aggiunta, permutazione'>
          <app>
            <rdg varSeq='1'><add hand='EM' place='supralinear,
marginleft'>al</add></rdg>
            <rdg varSeq='2'><del hand='EM' type='ripassato a
penna'>a</del>l</rdg>
            <rdg varSeq='3'><add hand='EM'
place='supralinear'>i</add>l</rdg>
            <rdg varSeq='4'><del hand='EM'
type='inchiostro'>i</del></rdg>
          </app>
        <rdg varSeq='3' type='aggiunta, permutazione'>
          <app>
            <rdg varSeq='1'><add hand='EM'
place='supralinear'>i</add></rdg>
            <rdg varSeq='2'><del hand='EM'
type='inchiostro'>i</del>l</rdg>
            <rdg varSeq='3'><add hand='EM'
place='supralinear'>a</add>l</rdg>
          </app>
        </rdg>
      </app>
    </rdg>
    <rdg wit='T, D' rend='italic'>nel</rdg>
    <rdg wit='Etutti, Mtuutti, O, At'>nel</rdg>
  </app>

```

TAVOLA 1

Una volta creato, l'apparato elettronico è stato 'letto' da un *browser*,<sup>9</sup> che ci ha permesso di vedere contemporaneamente, sullo schermo del computer:

a) il testo utilizzato come base, cioè i *Mottetti* contenuti ne *L'opera in versi*;

<sup>9</sup> La traduzione esatta del termine inglese è 'lettura rapida' ma noi potremmo definire il *browser* uno 'sfogliatore', cioè un programma di lettura, che riesce a rendere di nuovo leggibile un testo cui sia stata inserita una codifica.

- b) la sequenza delle variazioni di ogni singola parola o carattere, sapendo sempre a quali testimoni esse appartengano;
- c) l'anno di pubblicazione del testimone, in modo da ricostruire cronologicamente l'evoluzione di ogni singolo mottetto fino alla stesura definitiva;
- d) inoltre è possibile confrontare, sempre contemporaneamente, attraverso dei *links*, le varianti con le immagini digitalizzate dei testimoni, manoscritti o a stampa, in cui le varianti stesse sono state rilevate, onde consentire una immediata verifica del nostro lavoro da parte del nostro lettore/fruitor.

### *Analisi statistica*

Oltre alla costruzione dell'apparato elettronico, sono stati ottenuti in tal modo una serie di dati statistici, codificando ulteriormente il testo secondo le norme COCOA per il sistema TACT. Ovviamente i testimoni utilizzati per il processo di analisi sono gli stessi usati per la costruzione del documento SGML/TEI. Come già detto, ad ogni mottetto è stato assegnato un numero progressivo e accanto al numero è stata posta inoltre la sigla convenzionale del testimone da cui sono stati estratti i singoli mottetti.

Dunque ogni mottetto è stato identificato con il seguente formato:

7. L1

dove '7' è il numero del mottetto ed 'L1' è il testimone da cui il mottetto è stato estratto (che in questo caso è *La speranza di pure rivederti*, in "Letteratura", a. II, n. 1, Gennaio 1938, p. 68).

È stato identificato con un'etichetta ed un numero progressivo anche ogni verso. Per cui l'identificazione completa in ogni verso si presenta così:

<L 1.17.CP>Il fiore che ripete

dove 'L' sta per 'linea del verso', il numero a seguire, in questo caso '1', è il numero del verso, il successivo numero con la relativa sigla seguono la spiegazione data sopra. L'esempio appena riportato si riferisce quindi al verso n. 1 del mottetto n. 17 (*Il fiore che ripete*) pubblicato sul "Corriere Padano", nel 1938. L'ordine dato è stato dettato dalla struttura del sistema TACT che esige come prima etichetta per ogni linea di verso assolutamente la marca <L>, ed essendo tale marca l'abbreviazione di 'linea', non avremmo potuto collocare dopo questa nient'altro che il numero del verso.

I sintagmi apostrofati del tipo:

‘com’unghia’, ‘l’azzurro’, ‘all’opposta’

sono stati memorizzati ponendo uno spazio tipografico tra l’apostrofo e la parola che segue affinché, nello spoglio statistico, ogni singola parola che compone il sintagma fosse presa in considerazione come singola occorrenza.

Sono stati marcati anche quei termini o versi che sono variati nel corso del tempo, dalla prima stesura dei mottetti al 1980, anno dell’edizione critica a cura di Contini e Bettarini.<sup>10</sup>

Durante lo spoglio statistico relativo all’intera sezione e ai singoli mottetti, abbiamo avuto due possibilità, entrambe sfruttate, per l’analisi sulla punteggiatura usata da Eugenio Montale. Il sistema TACT ha bisogno di un ‘alfabeto’<sup>11</sup> (da noi costruito) con cui leggere il testo per cui in uno dei *files* utilizzati per la costruzione di tale ‘alfabeto’ non è stato inserito alcun segno di interpunzione.

Diverso è stato per lo studio delle varianti dove, per la codifica e lo spoglio delle stesse è stato necessario creare un *file* con un ‘alfabeto’ che contenesse i segni di interpunzione.

Il motivo è da ricercare nel fatto che nello spoglio statistico del testo base memorizzato per l’intera sezione, ci si è interessati della frequenza delle parole, delle occorrenze, delle concordanze, ecc. È stato invece interessante capire, in quale luogo Eugenio Montale abbia tolto o aggiunto una virgola, un punto e virgola e via dicendo, poiché anche questi segni rientrano nelle possibili varianti. Infatti per questo è stato istruito il programma in modo tale che il segno di interpunzione facesse ‘corpo unico’ con la parola immediatamente precedente e venisse quindi riconosciuto come un qualsiasi carattere. Per fare un esempio, a partire dall’edizione Mondadori del 1962, nel mottetto n. 3 (*Molti anni, e uno più duro sopra il lago*), verso 8, viene introdotta una virgola dopo la parola ‘fedeltà’ per cui il verso si presenta così:

<L 8.3>di fedeltà, immortale.

<sup>10</sup> Le raccolte pubblicate tra il 1962 (compresa) e il 1980 (esclusa *L’opera in versi*), non sono state prese in considerazione tranne il testimone ‘At’ (cfr. *supra*, nota 6).

<sup>11</sup> TACT è un sistema multilingue in quanto, utilizzando l’intero set dei caratteri ASCII, con esso si ha la possibilità di creare un *file* in cui avremo il profilo dell’ ‘alfabeto’ che si utilizzerà nel testo. In questo *file*, denominato \*.MKS (*MakeBase*), oltre l’alfabeto vengono definiti tutti i caratteri speciali, come ad esempio i segni diacritici, le etichette usate, e le referenze utilizzate nel testo. Una volta che il testo, memorizzato in un *file* di testo denominato \*.TXT, è stato marcato con le etichette scelte, il programma MAKEBASE lo converte in un *data base* trasformandolo in un *file* con estensione \*.TDB, utilizzabile con diversi programmi del sistema TACT.

mentre in tutte le edizioni precedenti il verso si presenta senza virgola:

<L 8.3>di fedeltà immortale.

per cui se volessimo sapere in quanti testimoni il termine 'fedeltà' è accompagnato dalla virgola e in quanti non lo è, dovremmo fare al programma una doppia interrogazione (*query*). La Tavola 2 che segue mostra la schermata della *Query Dialog Box* del programma USEBASE (sistema TACT) a cui abbiamo chiesto la collocazione del termine 'fedeltà' e dello stesso termine accompagnato dalla virgola: 'fedeltà,' (la seconda virgola che compare tra i due termini, serve al programma come separatore per poter effettuare più di una interrogazione nello stesso momento):

File	Select	Displays	Group	Help (F1)	93 K _
Query Dialog Box					
Query:					
fedeltà, . fedeltà					
Span Context: Unit: word					
Before: 5 After: 5					
For a list of Reference Names for "Unit", type "?" in the Unit field.					
Ignore Errors? Yes: _ No: X					
Manual Select? Yes: X No: _					
To choose option from Menu Bar, press the SPACEBAR					

TAVOLA 2: Finestra di dialogo per le *Query* del programma *UseBase* del sistema TACT.

I risultati ottenuti sono i seguenti (Tavola 3):

File	Select	Displays	Group	Help (F1)	61 K
<b>fedeltà (11)</b>					
8.3.C1		.....			fedeltà
8.3.GP1		.....			fedeltà
8.3.E1		.....			fedeltà
8.3.E2		.....			fedeltà
8.3.E3		.....			fedeltà
8.3.E4		.....			fedeltà
8.3.E5		.....			fedeltà
8.3.M1		.....			fedeltà
8.3.M2		.....			fedeltà
8.3.M3		.....			fedeltà
8.3.M4		.....			fedeltà
<b>fedeltà, (2)</b>					
8.3.At		.....			fedeltà,
8.3.O		.....			fedeltà,
KWIC Display					

TAVOLA 3: Programma *UseBase*, sistema TACT: risultati di una interrogazione (*query*)

Come possiamo vedere abbiamo ottenuto il termine 'fedeltà' e il termine 'fedeltà,' (con la virgola accanto) e le rispettive collocazioni in tutti i testimoni collazionati. Sarà la collocazione posta alla sinistra della schermata che ci darà indicazione del luogo cui queste occorrenze appartengono: come si è detto il primo numero corrisponde al numero del verso, il secondo al numero progressivo del mottetto, la lettera è la sigla del testimone. Con ulteriori istruzioni date al programma avremmo potuto ottenere anche la data di stesura, di pubblicazione, ecc.

Se non avessimo considerato i segni diacritici come caratteri legati alla parola, la stessa interrogazione ci avrebbe dato una segnalazione di errore da parte del programma in quanto quest'ultimo non avrebbe riconosciuto la prima virgola come elemento diacritico ma semplicemente come un separatore di termini, collocato in una errata posizione. Se invece chiedessimo esclusivamente la collocazione del termine 'fedeltà' (senza virgola accanto), otterremmo il seguente risultato (Tavola 4):

File	Select	Displays	Group	Help (F1)	61 K
<b>fedeltà (13)</b>					
8.3.C1		.....			fedeltà
8.3.GP1		.....			fedeltà
8.3.E1		.....			fedeltà
8.3.E2		.....			fedeltà
8.3.E3		.....			fedeltà
8.3.E4		.....			fedeltà
8.3.E5		.....			fedeltà
8.3.M1		.....			fedeltà
8.3.M2		.....			fedeltà
8.3.M3		.....			fedeltà
8.3.M4		.....			fedeltà
8.3.A1		.....			fedeltà
8.3.O		.....			fedeltà
KWIC Display					

TAVOLA 4: Programma *UseBase*, sistema TACT: risultati di una interrogazione (*query*)

dove abbiamo il termine richiesto, ma non sappiamo in quale dei testimoni Montale abbia utilizzato la virgola.

Il testo codificato con il linguaggio COCOA, per il sistema TACT, ci ha permesso di identificare e visualizzare tutte le varianti con e senza contesto e di fare una lista per ogni singola variante che ne mostri, ad esempio, il numero del verso, il numero del mottetto, e l'edizione in cui compare.

### Conclusioni

Il lavoro svolto ci ha fornito una serie di elenchi di numeri e di parole, liste di frequenza, lista numerica di *types* e di *tokens*, *index locorum*, concordanze, nonché tutti i versi che formano i *Mottetti* di Eugenio Montale, integralmente collazionati.

Attraverso la codifica con lo SGML/TEI, abbiamo dunque costruito un apparato variantistico dei *Mottetti* in linguaggio formato macchina (MRF), che ci ha permesso di scendere nel particolare e codificare anche la più piccola variante realizzando così un modello di testo comprendente tutti i testimoni che abbiamo collazionato. La nostra codifica si è estesa non solo alle varianti ma anche ai termini in corsivo, ai termini enfaticizzati, ai termini stranieri e inoltre è stata rispettata (per quanto possibile) la struttura tipografica del testo originale. Abbiamo così costruito una sorta di EDIC<sup>12</sup> di ogni documento che ci dà la possibilità, immediata, di visualizzarne le caratteristiche e soprattutto di poter gestire per qualsiasi ulteriore ricerca filologica le sue varianti.

Le strade da percorrere in merito allo sviluppo di una codifica delle varianti potevano essere due: nella prima, la più semplice, si sarebbe trattato di codificare i *Mottetti* tratti da *L'opera in versi*, insieme all'apparato magistralmente costruito da G. Contini e R. Bettarini; così tramite una delle tre metodiche previste dallo SGML/TEI, avremmo potuto costruire un apparato elettronico che rispondesse alla struttura di quello tradizionale già esistente.

La seconda strada, quella da noi percorsa, più difficile e lunga, ci ha dato la possibilità con la ricerca e la consultazione dei testimoni (dove ciò è stato possibile), di portare il nostro piccolo contributo all'ecdotica montaliana aggiungendo qualche variante a quelle già conosciute.

Attraverso il sistema TACT abbiamo praticamente 'smontato' il testo per dargli un ordine che ha, senza ombra di dubbio, la sua importanza. Probabilmente Montale nel comporre le sue poesie non avrebbe mai immaginato che un giorno i suoi versi sarebbero diventati tavole numeriche e percentuali, ma sicuramente, in qualche modo, anch'egli avrà contato le parole, le lettere, cercato anagrammi o figure retoriche. Non solo per costruire metricamente i suoi versi, ma anche per dare un'espressione più forte della sua sensibilità.

Osservando per esempio una qualsiasi delle liste di frequenza ottenute, il risultato del rapporto tra i *types* e i *tokens* ci ha evidenziato come in ogni mottetto queste due categorie siano quasi allo stesso livello numerico. Questo dato numerico conferma e dimostra una caratteristica dell'idioletto del Poeta già

<sup>12</sup> Edizione diplomatica interpretativa codificata.

intuito dai critici più avvertiti, e cioè che (al di là delle parti grammaticali quali congiunzioni, articoli, preposizioni, etc.) il lessico montaliano è vastissimo e ricercato.<sup>13</sup> Basti pensare ed osservare che nel mottetto n. 7 *Addii, fischi nel buio, cenni, tosse*, l'unico termine ripetuto due sole volte, è la congiunzione 'e'. Certamente è facile in un insieme di 38 vocaboli (o *tokens*) non ripetersi mai, assai meno facile (e del tutto caratteristico di Montale) è che questo accada in un insieme di 1209 parole! Eppure questo è quanto riscontriamo dalla analisi statistica dell'insieme dei *Mottetti*: ci sono pochissimi vocaboli ripetuti, per lo più, come si diceva, articoli, preposizioni, congiunzioni, etc.

Naturalmente l'interpretazione dei dati non è così semplicistica e va fatta con il testo alla mano perché, ad esempio, tra i *types* abbiamo degli omografi che non potremmo individuare se non avessimo l'intero testo davanti. Il *type* rappresenta dunque la 'classe' di un vocabolo, cioè troviamo in esso tutti quei vocaboli che sono graficamente identici tra loro. Questo fa sì che dal punto di vista della terminologia il lessico dei *Mottetti* (così come per qualsiasi altro testo) diventi ancora più vasto in quanto in uno stesso *type* possiamo trovare termini omografi ma con diversi significati, cioè in sostanza diverse parole.

Se queste nuove metodologie statistiche ci aiutano notevolmente nell'interpretazione, bisogna tuttavia fare attenzione che esse non vengano utilizzate in maniera arbitraria. Non si può certamente affermare, solo sulla base dei dati ricavati, che Montale sia (come vuole la tradizione critica) un poeta 'ermetico', ma si può sicuramente dire che egli sia senz'altro un poeta di difficile lettura. Si apre inoltre una strada assai promettente per affrontare i problemi di attribuzione di opere montaliane, un problema su cui ci ripromettiamo di tornare in altra sede. Valga questo esempio per affermare che il vecchio e il nuovo della critica letteraria debbano coesistere attivamente affinché si instauri «una sorta di rapporto dialettico tra testo, dati scaturiti dall'elaborazione informatica e ipotesi dello studioso, che realmente potrebbe aiutarci a disegnare un profilo del tutto nuovo dell'operazione di critica testuale».<sup>14</sup>

<sup>13</sup> La ricchezza del linguaggio in un testo è determinata dal rapporto tra le 'parole' e la 'classe' cui queste parole appartengono. Identificando la coppia 'parola/classe' con la corrispettiva coppia di termini inglesi 'token/type', possiamo dire che mentre il *token* può essere paragonato all' 'occorrenza', il *type* può essere definito come qualcosa a metà strada tra l' 'occorrenza' e il 'lemma'. Per fare un esempio, nei *Mottetti* la parola 'questo' appare due volte e costituisce un solo *type* ma due *tokens*; la parola 'questa' appare due volte e costituisce un altro diverso *type* e altri due *tokens*, e così via. Questo significa che il *token* viene catalogato in un *type*, senza dover essere riportato alla forma attestata nel dizionario, cioè al 'lemma' da cui deriva.

<sup>14</sup> G. Gigliozzi, *Il testo e il computer*, Milano, Bruno Mondadori, 1997, p. 255.

Lo studioso che costruirà il proprio modello, metterà allora a fuoco gli elementi misurabili del testo per riconoscerne le relazioni e costruire così un modello attendibile del testo da analizzare, naturalmente dichiarando e descrivendo puntigliosamente il lavoro svolto.

I due strumenti di lavoro, TACT ed SGML/TEI, ci hanno insegnato che le possibilità dei mezzi informatici sono molteplici e notevole è l'aiuto che l'umanista può trarre da tali mezzi se usati con correttezza e serietà. Le discussioni sul testo e la sua 'informatizzazione' sono appena iniziate così come si stanno sviluppando applicazioni concrete con la creazione di innumerevoli banche dati.

L'ultima considerazione che vorremmo fare riguarda la velocità con cui il progresso informatico avanza. Il rischio più grosso che si possa correre nel creare sempre nuovi programmi sempre più particolareggiati, che rendono un *software* appena costruito già obsoleto, è di perdere di vista l'oggetto reale della discussione: il testo. C'è il pericolo che questa discussione finisca per concentrarsi troppo sulla realizzazione di nuovi mezzi sempre più sofisticati e sullo sviluppo di nuove tecniche di codifica (che sono come abbiamo visto importantissime), per cui le due scienze, l'informatica e la filologia, debbono collaborare senza però invadere l'una il campo dell'altra. Solo così potremo avere degli ottimi strumenti che possano aiutarci nello scopo fondamentale, unico, del nostro lavoro: 'leggere il testo'.

Massimo Guerrieri