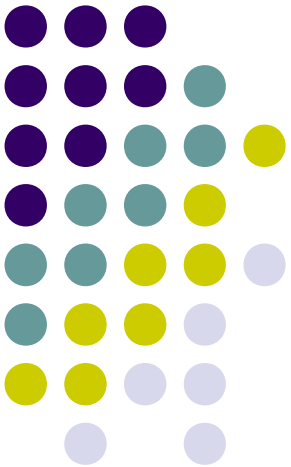
Metodi, modelli e strumenti
nell'analisi automatica del testo letterario

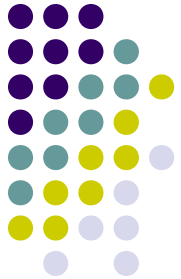
Nuove tecnologie, comunicazione, saperi umanistici

Roma, 21 Aprile 2005

Alessia Scacchi – Tiziana Mancinelli
Università di Roma "La Sapienza"



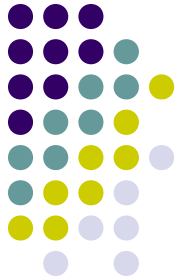
Critica letteraria e Analisi automatica



«La scelta che si compie quando, per svolgere un lavoro fin ora eseguito in maniera tradizionale, si decide d'usare un computer è d'affidare la nostra opera, o una parte di essa, a procedure algoritmiche»

GIUSEPPE GIGLIOZZI, *L'informatica, la didattica e il grillo-parlante*

Critica letteraria e Analisi automatica



La Dottoressa Calvin e…

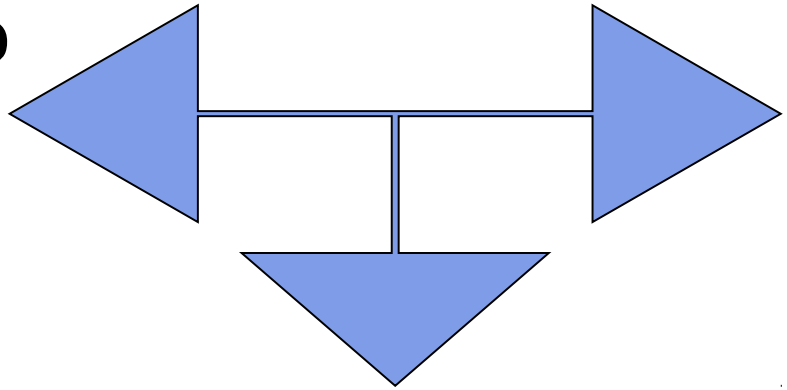
le tre leggi della robotica

- La Critica letteraria **supportata** dalla tecnica
- La tecnologia **applicata** all'interpretazione
- Interazione: **Humanities Computing**

Strumenti

Scelta del metodo interpretativo

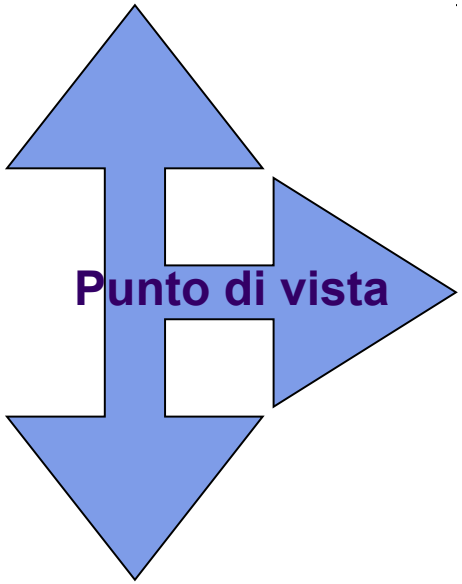
Strumentazione hardware e software



Modello ermeneutico

Modelli

- Soggetto: il critico

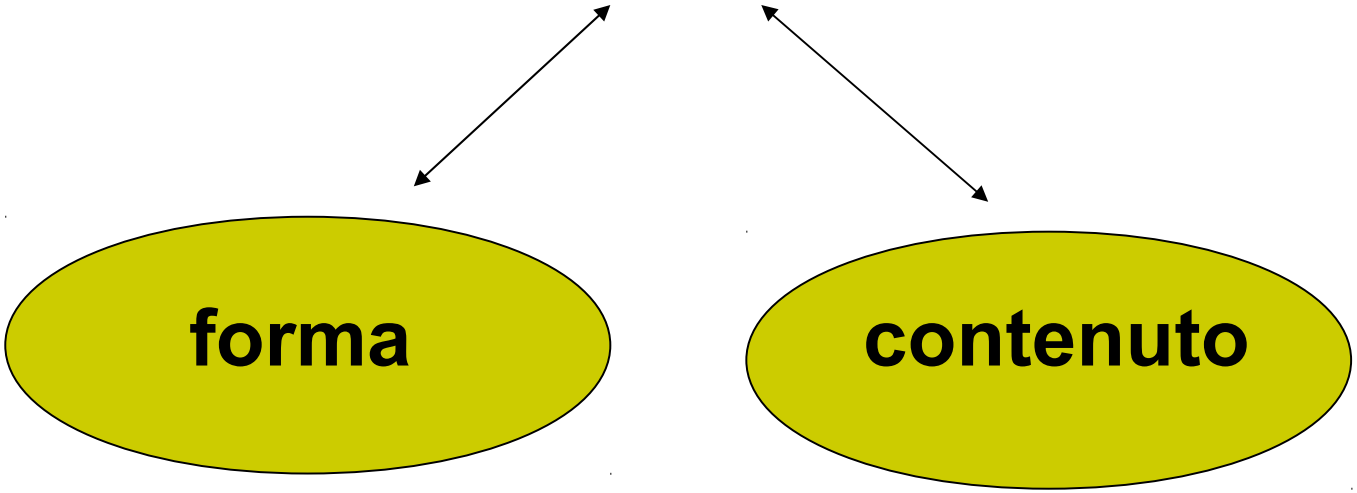


Modelli concettuali

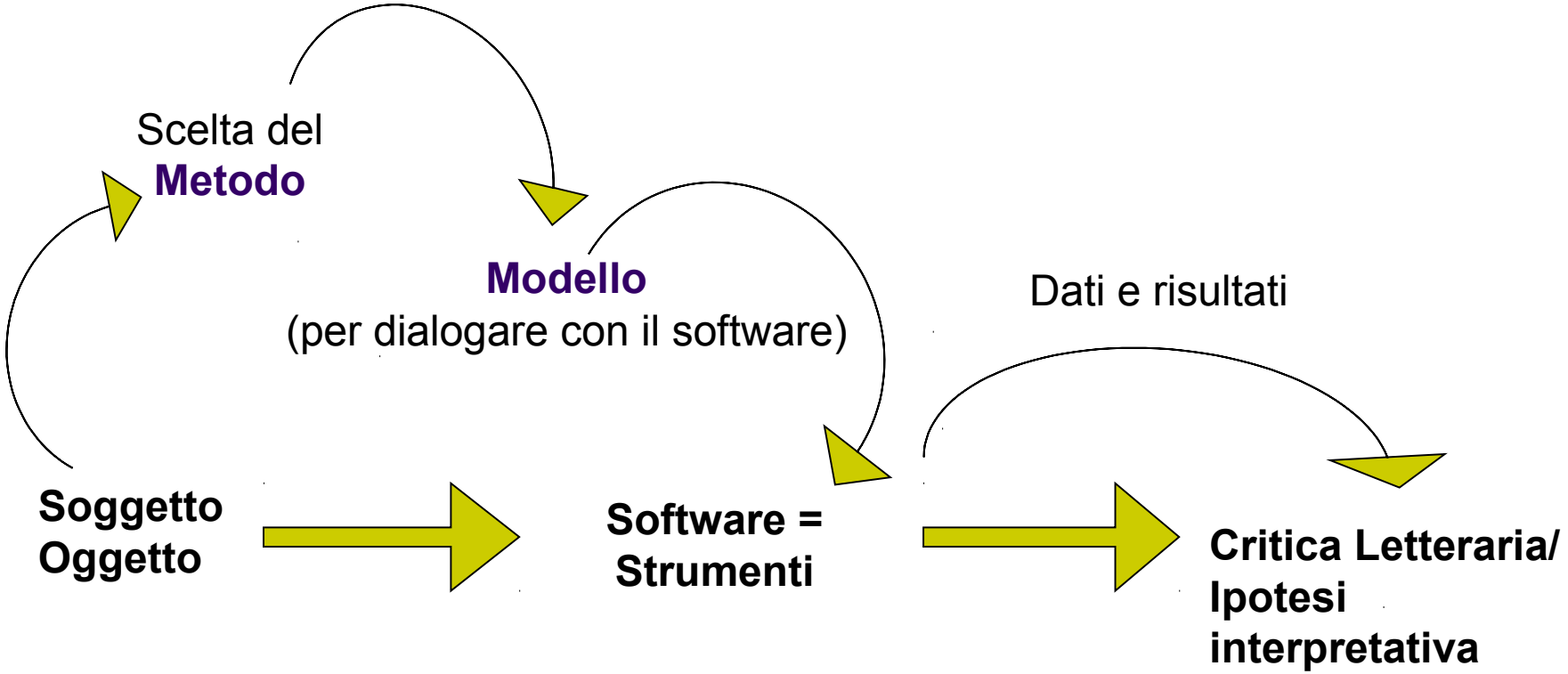
- Oggetto: il testo letterario

Metodo e strumentazione

Soggetto [in relazione con] Oggetto



Interazione metodi-modelli



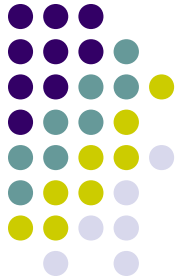
Modello epistemologico

Analisi del testo

*« Prendemmo a **spargere le carte** sul tavolo, **scoperte**, come per imparare a riconoscerle, e **dare loro** il giusto valore nei giochi, o **il vero significato nella lettura** del destino. […] Era qualcos'altro che vedevamo in quei tarocchi, qualcosa che non ci lasciava più staccare gli occhi dalle **tessere dorate di quel mosaico**»*

ITALO CALVINO, *Il castello dei destini incrociati*

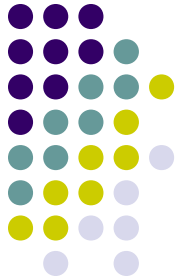
I. Analisi del testo



Fasi:

- Trasformazione del testo in una successione di caratteri alfa-numerici
- Lettura e correzione puntuale del testo (confronto cartaceo-digitale)
- Conta delle occorrenze dei termini
- **Marcatura di parti di testo secondo uno schema interpretativo precostituito**

II. Analisi del testo



- Il narrato è scomposto in singoli segmenti di significato interpretabili secondo:

  - Reiterazione
  - Posizionamento
  - Interazione reciproca

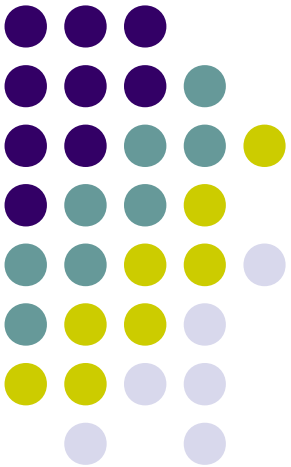
  L'analisi di questi elementi permette di rendere evidente la trama testuale, l'ordito sotteso alla stratificazione di significato propria della produzione letteraria

L’esempio: analisi automatica di un testo letterario

Estratto dalla tesi “Immaginario e corpo ne *L’agnese va a morire* di Renata Viganò. Un’analisi informatica del testo”

di

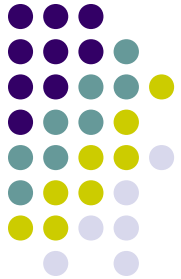
Alessia Scacchi



«l'enfasi si è spostata dal contenuto propositivo delle idee alla carica, alla qualità e al grado di intensità che esse esprimono»

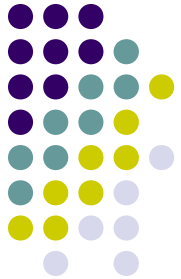
ROSI BRAIDOTTI, *In metamorfosi*

Finalità della ricerca



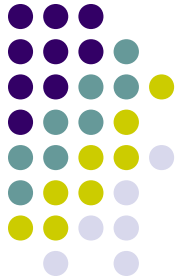
- Riconoscimento del valore di opere letterarie ideate e realizzate da donne
- Restituzione di senso all'opera letteraria e politica delle donne durante la Seconda Guerra mondiale
- Valorizzazione della scelta politica della narrazione e figurazione della guerra da un punto di vista interno all'azione

Modello: Scelta del Mark-up



- **<interp/>**
  - Il tag fondamentale previsto dalla TEI per annotare le parti del testo cui si intende applicare un'interpretazione particolare
    - www.tei-c.org/
- **<interp value='…'/>**
  - L'attributo indica a quale specifico fenomeno interpretativo si ritiene importante ricondurre la porzione di testo marcata
- **<interp ana='…'/>**
  - L'attributo associa l'elemento ad una o più caratteristiche cui si ritiene importante ricondurre l'analisi interpretativa

Modello:
I. Griglia interpretativa

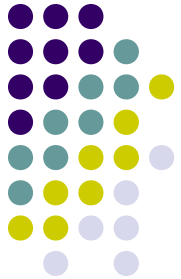


- Tag: **<interp/>**

L'interprete del testo sceglie per l'analisi le seguenti specifiche

- Attributi:
  - **value**='corpo' (per i riferimenti del testo al corpo in ogni sua parte)
    - **ana='**agnese**'** **(per i riferimenti alla protagonista)**
    - **ana='**donne**'** **(per i riferimenti a qualunque donna)**
    - **ana='**soldato**'** **(per i riferimenti ai soldati)**
    - **ana='**partigiani**'** **(per i riferimenti ai partigiani)**
    - **ana='**tedeschi**'** **(per i riferimenti ai tedeschi)**

Modello:
II. Griglia interpretativa

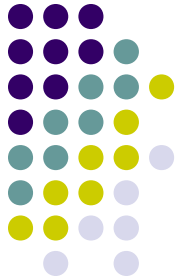


- Tag: **<interp/>**

L'interprete del testo sceglie per l'analisi le seguenti specifiche

- Attributi:
  - **value**='natura' (per i riferimenti del testo alla natura in ogni sua componente)
    - **ana='**acqua**'** **(per i riferimenti all'acqua)**
    - **ana='**terra**'** **(per i riferimenti alla terra)**
    - **ana='**fuoco**'** **(per i riferimenti al fuoco)**
    - **ana='**aria**'** **(per i riferimenti all'aria)**
    - **ana='**personificazione**'** **(per i riferimenti a personificazioni)**

Modello: III. Griglia interpretativa



- Tag: **<interp/>**

L'interprete del testo sceglie per l'analisi le seguenti specifiche

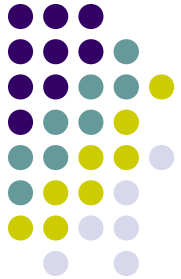
- Attributi:
  - **value**='sensazione' (per i riferimenti del testo alla natura in ogni sua componente)
    - **ana='**agnese**'** **(per i riferimenti alla protagonista)**
    - **ana='**donna**'** **(per i riferimenti a qualsiasi donna)**
    - **ana='**partigiani**'** **(per i riferimenti ai partigiani)**
    - **ana='**fascisti**'** **(per i riferimenti ai fascisti)** **in quanto i tedeschi nel testo non provano sensazioni**
    - **ana='**autore**'** **(per i riferimenti all'autore)**

Dynatext: Agnese.xml

Dati e risultati: I. Studio delle occorrenze

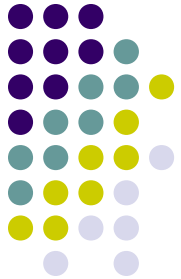
Numero d'ordine	Parole	Occorrenze (numeri assoluti)	Frequenze (percentuali)
1	E	2.016	2,67
2	LA	1.898	2,52
3	DI	1.882	2,5
4	IL	1.639	2,17
5	CHE	1.251	1,66
6	SI	1.160	1,54
7	UN	1.069	1,42
8	A	1.034	1,37
9	LE	1.034	1,37
10	NON	949	1,26
11	IN	869	1,15
12	I	857	1,14
13	PER	760	1,01
14	CON	654	0,87
15	UNA	621	0,82
16	ERA	566	0,75
17	**L'AGNESE**	**445**	**0,59**

Prima ipotesi interpretativa



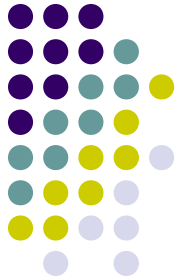
- Il femminile in stretta relazione con la poetica di Viganò:
  - <l'Agnese> prima parola significativa presente nel testo
  - Alta frequenza assoluta del sintagma
  - Distribuzione uniforme nell'intero testo

Dati e risultati: Studio dei *collocates*



Parole	Tot	L5	L3	L2	L1	Centre	R1	R2	R3	R4	R5
AGNESE	70	0	0	1	8	56	4	1	0	0	0
FACCIA	12	0	0	0	0	4	1	4	3	0	0
GROSSA	9	0	0	0	0	2	1	3	1	1	1
MANI	7	0	0	0	0	0	1	3	1	1	1
PESANTE	6	0	0	0	0	0	2	4	0	0	0
GAMBE	5	0	0	0	0	0	0	2	1	2	0
OCCHI	5	0	0	0	0	0	0	3	1	1	0
PIEDI	5	0	0	0	1	0	0	2	0	1	1
VOCE	7	3	1	0	0	0	0	0	1	2	0

Seconda ipotesi interpretativa

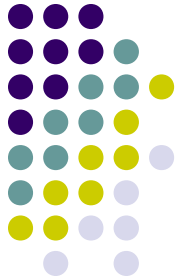


- Stretta relazione tra la rappresentazione del personaggio Agnese ed il suo corpo nelle singole parti
  - Prossimità tra il nome della protagonista e le annotazioni che riguardano il suo corpo
  - Alta frequenza assoluta dei termini che si riferiscono alle parti del corpo
  - Distribuzione omogenea nel testo

Dati e risultati: Studio delle occorrenze.2

Parola	Frequenza
Valle	133
Terra	116
Acqua	131
Argine	97
Neve	52
Fuoco	50
Campi	48
Sentiero	47
Vento	45
Canne	40
Cielo	37
Pioggia	35
Nebbia	34
Fiume	33
Ghiaccio	31
Fango	30

Terza ipotesi interpretativa



- Forte presenza della natura come parte integrante del tessuto testuale
  - Alta frequenza dei termini che indicano i quattro elementi
  - Forte legame tra la materia testuale e il contesto in cui è ambientata la vicenda (ambiente contadino)
  - Rappresentazione dei fenomeni atmosferici

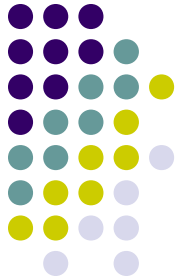
Dati e risultati: Studio delle frequenze

Termine	Frequenza assoluta
Partigian*	181
Tedesc*	350
Nero/a/i/e	30+32+13+14
Bianco/a/i/e	35+29+3+5
Casa/e	235+41

«L'Agnese restò sola, stranamente piccola, un mucchio di stracci neri sulla neve.»

RENATA VIGANÒ, *L'Agnese va a morire*

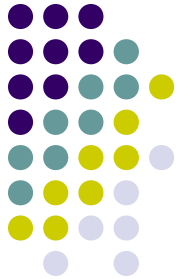
Modello e risultati: I. Ricerca complessa



- Value = **corpo**
- Ana =
  - agnese
  - donne
  - tedeschi
  - partigiani

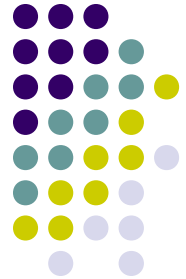
    - La ricerca individua le indicazioni relative al corpo
      - Della protagonista del romanzo
      - Delle donne presenti nel testo
      - Dei tedeschi
      - Dei partigiani

Modello e risultati: II.Ricerca complessa



- Value = **natura**
- Ana=
  - acqua
  - aria
  - fuoco
  - terra
    - La ricerca individua le parti del romanzo in cui la natura viene rappresentata con annotazioni relative
      - all'acqua
      - all'aria
      - al fuoco
      - Alla terra

Fine della storia… o inizio?



«Quello che lei sta affermando […] è che l'inadeguatezza di un robot deve essere controllata dalla versatilità e dall'intelligenza dell'uomo».

Isaac Asimov, *Il secondo libro dei robot.*