



RICERCHE DI MARKETING

ANALISI QUANTITATIVA  
DEI DATI QUALITATIVI

*Patrizia de Luca*





Ho chiesto a Lotaria se ha già letto alcuni miei libri che le avevo prestato. M'ha detto di no, perché qui non ha a disposizione un elaboratore elettronico.

M'ha spiegato che un elaboratore debitamente programmato può leggere un romanzo in pochi minuti e registrare la lista di tutti i vocaboli contenuti nel testo, in ordine di frequenza. – Posso così disporre subito d'una lettura già portata a termine, – dice Lotaria, – con un'economia di tempo inestimabile. Cos'è infatti la lettura d'un testo se non la registrazione di certe ricorrenze tematiche, di certe insistenze di forme e di significati? La lettura elettronica mi fornisce una lista delle frequenze, che mi basta scorrere per farmi un'idea dei problemi che il libro propone al mio studio critico. Naturalmente alle frequenze più alte sono registrate delle sfilze d'articoli, pronomi, particelle, ma non è là che soffermo la mia attenzione. Punto subito sulle parole più ricche di significato, che mi possono dare un'immagine del libro abbastanza precisa.

Lotaria m'ha portato alcuni romanzi trascritti elettronicamente sotto forma d'elenchi di vocaboli in ordine di frequenza. – In un romanzo tra le cinquantamila e le centomila parole, – m'ha detto, – le consiglio d'osservare subito i vocaboli che tornano una ventina di volte. Guardi qui. Parole che compaiono diciannove volte:

cinturone, comandante, denti, fai, han, insieme, ragno, risponde, sangue, sentinella, spari, subito, t', tua, visto, vita...

– Parole che compaiono diciotto volte:

basta, bello, berretto, finché, francese, mangiare, morto, nuovo, passa, patate, punto, quei, ragazzi, sera, vado, viene...

– Non ha già un'idea chiara di cosa si tratta? – dice Lotaria. –

Non c'è dubbio che è un romanzo di guerra, tutto azione, dalla scrittura secca, con una certa carica di violenza. Una narrazione tutta in superficie, si direbbe; ma per sincerarcene è sempre bene fare qualche sondaggio nella lista delle parole che ricorrono una volta sola, e non per questo sono meno importanti. Questa sequenza, per esempio:

sottana, sotterralo, sotterranei, sotterraneo, sotterrarla, sotterrato, sottili, sottobosco, sottomano, sottoproletari, sottoscala, sottoterra, sottovesti...

– No, non è un libro tutto in superficie come sembrava. Ci dev'essere qualcosa di nascosto; su questa traccia potrò indirizzare le mie ricerche.

Lotaria mi mostra un'altra serie d'elenchi. – Questo è un romanzo tutto diverso. Si vede subito. Guardi le parole che ricorrono una cinquantina di volte:

avuto, marito, poco, Riccardo, suo (51), cosa, davanti, ha, rispose, stata, stazione (48), appena, camera, Mario, qualche, tutti, volte (47), andò, cui, mattina, pareva (46), doveva (45), avesse, fino, mano, senti (43), anni, Cecina, chi, Delia, mani, ragazza, sei, sera (42), finestra, poteva, quasi, sola, tornò, uomo (41), me, voleva (40), vita (39)...

– Cosa glie ne pare? Narrazione intimista, sentimenti sottili, appena accennati, un ambiente modesto, la vita di tutti i giorni in provincia... Per controprova, preleviamo un campione di parole che ricorrono una sola volta:

infreddolito, ingannata, ingegnato, ingegnere, ingelosire, ingenua, inghiottì, inghiottita, inghiottiva, inginocchiarsi, ingiù, ingiustizia, ingrandiva, ingrassare...

– E così, già ci siamo resi conto dell'atmosfera, degli stati d'animo, dello sfondo sociale... Possiamo passare a un terzo libro:

andò, capelli, conto, corpo, Dio, secondo, soldi, soprattutto, vol (39), farina, pioggia, provviste, qualcuno, ragione, sera, stare, Vincenzo, vino (38), dolce, dunque, gambe, morte, sue, uova, verde (36), avremmo, bambini, beh, bianco, capo, fanno, giornata, macchina, neri, persino, petto, rimasi, sta, stoffe (35)...

– Qui direi che siamo di fronte a una storia corposa, sanguigna, tutta sul sodo, un po' brusca, con una sensualità diretta, senza raffinatezze, un erotismo popolaresco. Passiamo anche qui alla lista delle parole con frequenza uno. Ecco, per esempio:

verdure, vergini, vergognai, vergognandosi, vergognare, vergognarti, vergognasse, vergognata, vergognava, vergogne, vergogneremmo, vergogni, vergogno, verificarsi, vermut... .

– Ha visto? Questo è un senso di colpa bello e buono! Un indizio prezioso: l'indagine critica può partire di lì, proporre le sue ipotesi di lavoro... Cosa le dicevo? Non è un sistema rapido ed efficace?



L'idea che Lotaria legga i miei libri a questo modo mi crea dei problemi. Adesso ogni parola che scrivo la vedo già centrifugata dal cervello elettronico, disposta nella graduatoria delle frequenze, vicino ad altre parole che non so quali possano essere, e mi domando quante volte l'ho usata, sento la responsabilità dello scrivere che pesa tutta su quelle sillabe isolate, provo a immaginarmi quali conclusioni si possano trarre dal fatto che ho usato una volta o cinquanta volte quella parola. Forse sarà meglio che la cancelli... Ma qualsiasi altra parola provi a sostituirla, mi sembra che non resista alla prova... Forse anziché un libro potrei scrivere degli elenchi di parole, in ordine alfabetico, una frana di parole isolate in cui si esprima quella verità che ancora non conosco, e dalle quali l'elaboratore, capovolgendo il proprio programma, ricavi il libro, il mio libro.



1979

# Analisi del contenuto di un testo (Analisi testuale)

- Il contenuto di un testo può essere elaborato in modo scientifico, utilizzando i metodi e gli strumenti messi a disposizione dalle discipline quantitative (matematica e statistica).
- Analisi tradizionale del contenuto di Harold Lasswell (1949), uno dei primi tentativi di semantica quantitativa
  - *Si limita a contare le occorrenze dei termini nel testo.*
  - *Il significato della parola con la frequenza assoluta più elevata determina il senso generale del discorso.*
  - *Ma le parole hanno diversi significati: come scegliere quello giusto?*
  - *Il singolo termine non ci permette di saperlo: è necessario analizzare l'intera frase in cui si trova la parola, ossia le altre parole o co-occorrenze.*
- I programmi informatici, sviluppati a partire dalla fine degli anni Sessanta del XX secolo, hanno permesso l'applicazione di nuove tecniche di analisi.

# Analisi quantitativa del contenuto

- Punta al computo dei dati e fornisce un'analisi descrittiva di tipo inferenziale
- Si basa sulla misura della ricorrenza all'interno del testo delle unità di analisi: assume la distribuzione delle **frequenze** come base per i processi di inferenza
- Le unità di analisi stabilite sulla base degli obiettivi del ricercatore sono raggruppate in **categorie**, definite a loro volta in **unità di classificazione**.
- Le unità di analisi e unità di classificazione sono analizzate tramite diverse tipologie di analisi quantitativa:
  - *l'analisi delle frequenze*
  - *l'analisi delle contingenze*
  - *l'analisi delle corrispondenze lessicali*

# Interpretazione

- La griglia concettuale rappresenta la base per l'interpretazione
- Possibilità di comparare agevolmente i contenuti e di cogliere le diverse modalità di costruzione dei discorsi intorno a ogni tema
- Tale comparazione rappresenta *“un confronto tra contesti, cioè un confronto del complesso delle considerazioni emerse su un dato argomento e sui collegamenti con altri argomenti”*
- La griglia consente di differenziare contenuti trasversali alle unità di contesto (ad esempio, interviste) e specifici a determinati contesti
- Disponendo della griglia è possibile infatti ricondurre le diverse modalità di articolazione dei contenuti a caratteristiche specifiche, ad esempio, dell'intervistato e del contesto in cui opera
  - *Ciò permette al ricercatore di conseguire una visione del contesto completa e funzionale alla realizzazione dello specifico obiettivo conoscitivo dell'indagine*
- Al fine di assicurare la verificabilità dell'analisi, ossia poter garantire che altri ricercatori giungano a medesimi risultati in base ai documenti a disposizione e ai dati grezzi, più ricercatori sono generalmente coinvolti sia nella fase di codifica, sia nella fase di interpretazione

# L'uso dei software per l'analisi del contenuto

- I **software** oggi disponibili aiutano nel codificare contenuto di testi anche molto ampi
- I software aiutano nella codifica del materiale ma non nell'interpretazione
- L'unità di analisi è tipicamente la parola in quanto unità minima di senso ottenibile dalla massima disgregazione dei testi
- Calcolo di concordanze, co-occorrenze, frasi tipiche
- Problema dell'ambiguità della parola rende necessaria una verifica e un miglioramento delle attribuzioni categoriali effettuate dal software

# Quando l'analisi quantitativa del contenuto non è appropriata

- Se le domande di ricerca sono ampie, esplorative, variabili e lavorare con categorie risulterebbe restrittivo
- Se si vuole utilizzare un approccio olistico, e non un'analisi che procede per tappe pianificate

# Analisi delle frequenze

- Si assume che la ricorrenza della categoria (o unità di classificazione) nel testo dimostri l'importanza che essa riveste al suo interno.
- Obiettivo: rilevare i temi più importanti all'interno di un testo
- Metodo: si misura la frequenza con cui una certa unità (parola, concetto, ecc.) è presente nel testo relativamente alle altre.



# Analisi delle contingenze

- Osgood (1959): procedura che presuppone la classificazione delle unità di analisi, per determinare la presenza contemporanea di una o più unità di categorie diverse in uno stesso testo o segmento di testo assunto come unità di contesto
- Quante volte le unità di analisi appartenenti ad una certa categoria si presentano insieme a quelle appartenenti a un'altra, nella stessa unità di contesto (nella stessa porzione di testo o nel testo intero)?
- E' possibile così rilevare le associazioni fatte nel testo tra le diverse categorie (si possono definire meglio opinioni, atteggiamenti, ecc.)
- Utile per rilevare tratti di personalità, atteggiamenti, rappresentazioni sociali, intenzioni latenti, ...
- Tecnicamente va costruita la matrice delle contingenze (tabella a doppia entrata)

## Esempio di tabella a doppia entrata per analisi delle contingenze

<i>capelli\occhi</i>	<i>chiari</i>	<i>scuri</i>	<i>totale</i>
<i>chiari</i>	7	18	25
<i>scuri</i>	13	62	75
<i>totale</i>	20	80	100

# Le co-occorrenze

(o concomitanze o contingenze di contenuto)

- Sono le altre parole che accompagnano un termine in un testo
- Quando si parla di una particolare accezione del termine, alcune parole si presentano nel discorso con una frequenza maggiore rispetto alle altre

# Le co-occorrenze: esempio

(o concomitanze o contingenze di contenuto)

Si consideri il seguente esempio, relativo al termine «pesca»:

- A) La pesca sportiva è uno sport, può essere praticato sia in corsi d'acqua dolce che in mare.
- B) Chi pratica la pesca sportiva ama l'acqua e il mare, cattura il pesce solo per sfida.
- C) La pesca è un'attività ittica molto importante, il pesce è uno degli alimenti principali.
- D) I premi in palio della pesca di beneficenza sono stati donati dal consorzio.
- E) L'estrazione dei numeri vincenti della pesca di beneficenza avverrà venerdì.
- F) Il ricavo della vendita dei biglietti della pesca andrà in beneficenza.

[www.seosemantica.it](http://www.seosemantica.it)

# Le co-occorrenze: esempio

- Parlando di «pesca» come attività ittica, può capitare di trovare delle parole concomitanti come «pesce», «pescatore», «rete», ...
- Se per pesca si intende l'estrazione di beneficenza, nel discorso compaiono altre parole concomitanti, come «estrazione», «biglietto», ...
- Le parole concomitanti capitano frequentemente nei discorsi a seconda del contesto, insieme al termine principale di riferimento (pesca, nell'esempio), e quindi sono dette contingenze o co-occorrenze
- Analizzando le contingenze (co-occorrenze) è possibile capire il significato della parola principale di riferimento (pesca), eliminando tutte le altre accezioni.

# L'analisi delle co-occorrenze:

## obiettivo e procedura

- Il ricercatore individua N documenti (ad esempio, i testi delle interviste) dove è presente il termine da analizzare, ad esempio «pesca»
- Elimina da ogni documento le parole vuote (articoli, preposizioni, ecc.) per ridurre il numero dei termini alle sole parole piene, cioè quelle che hanno almeno un significato proprio
- Cerca inoltre di raggruppare i termini per non presentare due volte lo stesso termine al maschile o al femminile, al singolare o al plurale, ecc.
- Il ricercatore costruisce la tabella delle contingenze mettendo in colonna tutte le parole individuate nei documenti, evitando le duplicazioni.
  - *Ad esempio, se la parola "sportiva" compare cinque volte in tre documenti, viene comunque associata a una sola colonna nella tabella.*

(Fonte: adattamento da <http://www.andreaminini.com/semantica/analisi-delle-contingenze-del-contenuto>, 9-11-2016)

# Esempio

~~La pesca sportiva è uno sport, può essere praticato sia in corsi d'acqua dolce che in mare.~~

~~Chi pratica la pesca sportiva ama l'acqua e il mare, cattura il pesce solo per sfida.~~

~~La pesca è un'attività ittica molto importante, il pesce è uno degli alimenti principali.~~

~~I premi in palio della pesca di beneficenza sono stati donati dal consorzio.~~

~~L'estrazione dei numeri vincenti della pesca di beneficenza avverrà venerdì.~~

~~Il ricavo della vendita dei biglietti della pesca andrà in beneficenza.~~

## parole piene

pesca	premi
sportiva	palio
è	beneficenza
può	sono
essere	stati
praticato	donati
corsi	consorzio
acqua	estrazione
dolce	numeri
mare	vincenti
pratica	avverrà
ama	venerdì
cattura	ricavo
pesce	vendita
sfida	biglietti
attività	andrà
ittica	
molto	
importante	
alimenti	
principali	

[www.seosemantica.it](http://www.seosemantica.it)

## parole piene

pesca	premi
sportiva	palio
<b>è</b>	beneficenza
può	<b>sono</b>
<b>essere</b>	stati
praticato	donati
corsi	consorzio
acqua	estrazione
dolce	numeri
mare	vincenti
pratica	avverrà
ama	venerdì
cattura	ricavo
pesce	vendita
sfida	biglietti
attività	andrà
ittica	
molto	
importante	
alimenti	
principali	

Essere

sportivo

sportivo

sportivi

sportiva

[www.seosemantica.it](http://www.seosemantica.it)



- Il passo successivo consiste nel mettere al vaglio tutte le frasi o le proposizioni dei documenti, inserendole nelle righe della tabella.
- Ogni riga della tabella è associata a una proposizione.
- Per ciascuna proposizione (riga) il ricercatore verifica quali termini in colonna sono presenti.
- Quando un termine compare in una proposizione, il ricercatore inserisce il numero uno (1) nella relativa cella della tabella, quella individuata dall'incrocio tra la colonna I del termine e la riga J della proposizione. Viceversa, quando il termine (colonna) non è presente nella proposizione (riga), viene inserito il numero zero (0) nella cella.
- Questo processo, detto operativizzazione, permette di trasformare le frasi in vettori composti da sequenze di zero e uno. I vettori sono molto più facili da confrontare rispetto alla forma letterale delle frasi.

## Tabella delle occorrenze: esempio

PROPOSIZIONI	pesca	sportiva	sport	essere	acqua	pesce	premio	palio	beneficenza	estrazione	biglietto	vincente
La pesca sportiva è uno sport, può essere praticato sia in corsi d'acqua dolce che in mare.	1	1	1	1	1	0	0	0	0	0	0	0
Chi pratica la pesca sportiva ama l'acqua e il mare, cattura il pesce solo per sfida.	1	1	0	0	1	1	0	0	0	0	0	0
La pesca è un'attività ittica molto importante, il pesce è uno degli alimenti principali.	1	0	0	1	1	1	0	0	0	0	0	0
I premi in palio della pesca di beneficenza sono stati donati dal consorzio.	1	0	0	1	0	0	1	1	1	0	0	0
L'estrazione dei numeri vincenti della pesca di beneficenza avverrà venerdì.	1	0	0	0	0	0	0	0	1	1	0	1
Il ricavo della vendita dei biglietti della pesca andrà in beneficenza.	1	0	0	0	0	0	0	0	1	0	1	0

www.seosemantica.it

## Dalla tabella delle occorrenze alla tabella delle associazioni semantiche

- Una volta completata l'operativizzazione, il ricercatore costruisce una matrice quadrata ponendo nelle colonne e nelle righe i termini.
- Nelle celle è, invece, indicato il numero dei documenti in cui compare la combinazione dei termini.
- In questo modo, il ricercatore ha ottenuto una tabella delle associazioni semantiche.
- Questa tabella ci permette di ottenere dei cluster, ossia dei gruppi di termini, in cui il termine oggetto di studio ("pesca" ) ha un significato piuttosto che un altro.
- Nell'esempio precedente le associazioni semantiche si concentrano in due cluster ( raggruppamenti ), ognuno dei quali mostra un significato diverso della parola "pesca". Nel primo cluster blu le co-occorrenze indicano l'atto di catturare il pesce. Nel secondo cluster rosso, invece, mostrano l'atto di estrarre dei numeri.

## Tabella quadrata delle occorrenze (con associazioni semantiche): esempio

### MATRICE QUADRATA DELLE OCCORRENZE

### PAROLE PIENE

PAROLE PIENE	pesca	sportiva	sport	essere	acqua	pesce	premio	palio	beneficenza	estrazione	biglietto	vincente
pesca	X	2	1	3	3	2	1	1	3	1	1	1
sportiva	2	X	1	1	2	1	0	0	0	0	0	0
sport	1	1	X	1	1	0	0	0	0	0	0	0
essere	3	1	1	X	2	1	1	1	1	0	0	0
acqua	3	2	1	2	X	2	0	0	0	0	0	0
pesce	2	1	0	1	2	X	0	0	0	0	0	0
premio	1	1	0	1	0	0	X	1	3	0	0	0
palio	1	0	0	1	0	0	1	X	1	0	0	0
beneficenza	3	0	0	1	0	0	3	1	X	1	1	1
estrazione	1	0	0	0	0	0	0	0	1	X	0	1
biglietto	1	0	0	0	0	0	0	0	1	0	X	0
vincente	1	0	0	0	0	0	0	0	1	1	0	X

[www.seosemantica.it](http://www.seosemantica.it)

MATRICE QUADRATA  
DELLE OCCORRENZE

PAROLE PIENE

PAROLE PIENE	pesca	sportiva	sport	essere	acqua	pesce	premio	palio	beneficenza	estrazione	biglietto	vincente
pesca	X	2	1	3	3	2	1	1	3	1	1	1
sportiva	2	X	1	1	2	1	0	0	0	0	0	0
sport	1	1	X	1	1	0	0	0	0	0	0	0
essere	3	1	1	X	2	1	1	1	1	0	0	0
acqua	3	2	1	2	X	2	0	0	0	0	0	0
pesce	2	1	0	1	2	X	0	0	0	0	0	0
premio	1	1	0	1	0	0	X	1	3	0	0	0
palio	1	0	0	1	0	0	1	X	1	0	0	0
beneficenza	3	0	0	1	0	0	3	1	X	1	1	1
estrazione	1	0	0	0	0	0	0	0	1	X	0	1
biglietto	1	0	0	0	0	0	0	0	1	0	X	0
vincente	1	0	0	0	0	0	0	0	1	1	0	X

pesca come attività ittica

pesca come estrazione

www.seosemantica.it

# L'analisi delle corrispondenze lessicali: cenni e rinvii

- Metodo elaborato dal francese Benzécri negli anni '80, prendendo spunto dall'analisi delle co-occorrenze di Osgood
- Conosciuta anche con l'acronimo inglese LCA (Lexical Correspondence Analysis )
- Applicazione ai testi dell'analisi fattoriale delle corrispondenze.

Fonte: adattamento da <http://www.okpedia.it/analisi-delle-corrispondenze-lessicali>

# ESERCITAZIONE

Atlas.ti