L'analisi testuale e l'analisi delle corrispondenze lessicali

Maria Paola Piccini
mariapaola.piccini@uniroma1.it
mariapaolapiccini@virgilio.it
Università Pontificia Salesiana –
Facoltà di Scienze della Comunicazione Sociale

L’analisi statistica di dati testuali

«Ho chiesto a Lotaria se ha già letto alcuni miei libri che le avevo prestato. Mi 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. “Cos’è infatti la lettura d’un testo se non la registrazione di certe ricorrenze tematiche, di certe insistenze di forme e di significati?” (…) 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…»

Italo Calvino, *Se una notte d’inverno un viaggiatore* (1979)

Quali testi analizzare?

I diversi approcci di ricerca, standard e non, tra i diversi elementi in comune, hanno la cosiddetta **cumulazione del dato**, la cumulazione di una gran mole di **materiale testuale**, che costituisce una fonte di informazioni sicuramente preziose, che deve però essere in qualche modo "governata".

Con l'analisi testuale è possibile esplorare sistematicamente, rapidamente e, talvolta in modo semi-automatico, la struttura dei testi, anche molto ampi e difficilmente esplorabili in altro modo, con il vantaggio di poter tornare in ogni momento ai testi originari.

Quali testi analizzare?

Protocolli di interviste, trascrizioni di focus group, risposte a domande aperte di questionari, storie di vita, documenti, verbali, note di osservazione, etc.

Ma anche conversazioni sui newsgroup, blog, forum etc.

L’analisi testuale

- È importante sottolineare che le procedure di analisi testuale non si limitano semplicemente al **conteggio** delle singole parole o forme del corpus, ma con esse è possibile approfondire i **contenuti** in esso presenti, per mezzo di operazioni di inventario, ricerca (*text retrieval*), selezione e classificazione dei testi o parti di testo, fino alla rappresentazione grafica delle forme su un piano delimitato da due assi fattoriali, così da individuare dimensioni di senso latenti, sottese ai dati testuali stessi.

L’analisi testuale

- Le diverse tecniche di analisi testuale, attraverso il supporto di specifici *software*, rispondono all’esigenza di accostarsi a campi d’indagine complessi e consentono l’esplorazione, la descrizione e l’analisi di corpora testuali anche molto estesi e/o poco strutturati (della Ratta-Rinaldi, 2007).
- L’analisi testuale permette di ampliare la visione d’insieme del problema oggetto di studio facendo emergere dimensioni semantiche e tematiche talvolta inattese, soggiacenti agli stessi dati testuali e mettendo in luce il punto di vista dei produttori dei testi analizzati.

L’analisi con le tecniche di statistica testuale

- Poter utilizzare tecniche statistiche di elaborazione dei dati testuali costituisce senza dubbio un vantaggio per il ricercatore, che dovrà tuttavia guardarsi dal rischio eventuale di **decontestualizzazione** delle parole e di possibili **eccessi di automatismo**. Nessun approccio di analisi sostituisce il compito di riflettere sul significato dei dati e “nessun automatismo può supplire da solo alla conoscenza tacita che si esprime nel con-testo e nell’extra-testo” (Giuliano, 2004, p. 122).
- In questa prospettiva, con l’analisi testuale si cerca di individuare uno **schema interpretativo** sotteso alla lettura diretta del testo (Giuliano, 2004).

L’analisi con le tecniche di statistica testuale

- Il ricorso alle tecniche di analisi testuale è considerato come una delle **strategie ponte**, terreno d’incontro e integrazione per eccellenza tra metodi quantitativi e qualitativi, grazie alla possibilità di coniugare la necessità di produrre **studi di tipo empirico controllabili pubblicamente**, con la **ricchezza interpretativa** (della Ratta-Rinaldi, 2002).

Diversi approcci all'analisi dei testi

Due approcci prevalenti all'analisi dei testi, che dipendono dalle dimensioni del testo e dagli obiettivi dell'analisi

	ermeneutico	lessicometrico
Vantaggi	- analisi contestuale - analisi relazioni tra concetti - costruzione mappe concettuali - visualizzazione del testo	- Testi molto ampi - confronti tra diverse parti del testo - ricorso a tecniche statistiche e a fonti linguistiche esterne - possibilità di ritorno al testo
Svantaggi	- sconsigliata per testi ampi - labour intensive - soggettività codifica operatori - difficile ispezionabilità procedure	- decontestualizzazione parole - si lavora per parole e non per concetti - difficoltà a cogliere ambiguità linguaggio - possibili eccessi di automatismo

I principali pacchetti statistici per l'analisi testuale lessicometrica

Prima descrizione ed esplorazione del corpus:

- ✓ Lexico3

(http://www.cavi.univ-paris3.fr/ilpga/syled/index.htm),

- ✓ MonoconcPro (http://www.athel.com/mono.html);

Analisi del linguaggio e trattamento del testo:

Taltac (http://www.taltac.it)

Tlab (http://www.tlab.it/it/presentation.php)

I principali pacchetti statistici per l'analisi testuale lessicometrica

Analisi multidimensionali:

- Spad
  http://www.coheris.com/fr/page/produits/Spad.html
- Alceste
  http://www.image-zafar.com/en/alceste-software
- Sphink http://scolari.co.uk

Il software Spad

Spad è un *software* per l'analisi semi-automatica dei testi che si avvale di tecniche statistiche e lessicali basate sull'analisi delle parole e delle loro relazioni all'interno del testo.

È un programma particolarmente indicato per lo studio e l'esplorazione sistematica delle **dimensioni di senso** presenti in un corpus testuale, infatti, permette di individuare in modo **semi-automatico** i contenuti o gli argomenti principali del testo oggetto di studio,

assumendo la **frequenza** delle parole e/o delle categorie tematiche come **indicatore** della **rilevanza** di ciascun **tema** (della Ratta-Rinaldi, 2007).

Il software Spad

Oltre alla descrizione dei **contenuti** di un corpus, con Spad, è possibile far interagire il testo con le variabili disponibili sui **produttori del testo** stesso,

in questo modo si possono identificare eventuali **differenze peculiari** nell'uso delle parole o nella scelta degli argomenti, tra i diversi tipi di produttori dei testi.

Glossario

Un **corpus** di testi è una qualsiasi raccolta di frammenti testuali fra loro **confrontabili** (documenti, verbali, domande aperte, resoconti di focus group, interviste libere, raccolta di lettere o canzoni, etc.). Il corpus è l'insieme dei testi sui quali si vuole e si deve effettuare l'analisi, è un insieme ragionato di testi che corrispondono ad un obiettivo, allo scopo per cui verrà analizzato. "Per corpus s'intende un qualsiasi insieme di testi, fra loro confrontabili sotto un qualche punto di interesse" (Bolasco, 1999, p. 182).

Occorrenza: Ogni parola che compare in un testo.

Vocabolario: insieme di parole diverse del corpus. Può essere espresso o in forme grafiche (le parole così come compaiono nel corpus), o per lemmi, cioè le forme presenti nei dizionari

Dimensione: Il numero totale di occorrenze del corpus

Condizioni per l'analisi

- Comparabilità dei testi
- Dimensioni
- Rilevanza
- Possibilità di associare una o più variabili ai testi (notizie sugli autori, data di produzione etc.)

Comparabilità e Dimensioni dei testi

I testi dovrebbero essere **comparabili** fra loro per struttura, autori, destinatari, dimensioni.

Generalmente, è necessario disporre di testi sufficientemente lunghi:

un corpus è considerato **piccolo** quando non supera le **15.000** occorrenze,

medio quando raggiunge le **45.000** occorrenze e

medio-grande quando supera le **100.000** (Bolasco, 1999).

Dimensioni

- L’analisi testuale e, più specificamente, l’analisi lessicometrica, si prestano tendenzialmente meglio a **corpora di grandi dimensioni**. Testi di piccole dimensioni possono essere comunque sottoposti ad analisi, ma con potenzialità ridotte e, soprattutto, con una ridotta portata e solidità dei risultati.
- La **ricchezza** dei testi può supplire a dimensioni ridotte (es. testi registrati piuttosto che risposte autocompilate)

Rilevanza

- I testi da analizzare devono essere progettati all'interno di un percorso di analisi ragionato e non analizzati solo perché disponibili!
- La resa migliore dell'analisi testuale si ottiene associandola ad altri strumenti di ricerca e di analisi.

La dimensione tematica: una strategia di analisi

1. Acquisizione e normalizzazione del testo;
2. Assegnazioni di chiavi ai testi per la comparazione tra tipi di testi;
3. Studio del **vocabolario** (parole **piene** e parole **chiave**);
4. Selezione e classificazione dei **segmenti ripetuti** significativi;
5. Analisi dei contesti delle parole finalizzata alla individuazione dei temi/discorsi portanti del testo;
6. Classificazione delle parole **piene** e dei **segmenti ripetuti significativi** in categorie;
7. Calcolo delle parole e/o delle categorie **specifiche** (o caratteristiche) per tutte le chiavi associate ai testi;
8. Analisi delle corrispondenze lessicali.

1. Costruzione e pre-trattamento

- Se si trascrive un testo si possono decidere accortezze (evitare maiuscole, usare abbreviazioni coerenti etc.)

- Se si lavora su un testo "scansionato" (o scaricato dalla rete) è necessario lavorare sulla **normalizzazione** del testo

Normalizzazione

Generalmente, prima di procedere nell'analisi, è necessario sottoporre il corpus alla **normalizzazione**, che consiste nell'**omogeneizzazione delle grafie** utilizzate e che viene effettuata al fine di eliminare le possibili fonti di sdoppiamento del dato testuale, abbassando le maiuscole oppure uniformando la grafia dei nomi propri, delle sigle, dei numeri e delle date, che solitamente comportano una forte variabilità. Infatti, il *software* riconosce le parole presenti nel testo come **successioni di caratteri** dell'alfabeto comprese **fra spazi vuoti** o fra i più comuni segni di punteggiatura, detti **separatori**.

Normalizzazione

Di conseguenza, in assenza dell'applicazione della normalizzazione del testo, il *software* interpreterebbe come parole diverse, ad esempio, la stessa parola scritta con l'iniziale maiuscola e con l'iniziale minuscola.

Tuttavia, per il ricercatore a volte può essere utile mantenere la distinzione fra parole scritte con l'iniziale maiuscola o minuscola, ad esempio, per conservare la differenziazione fra la parola **chiesa** quando sta ad indicare il **luogo fisico**, e la parola **Chiesa** quando invece sta ad indicare **l'istituzione religiosa**.

Lemmatizzazione

Oltre alla normalizzazione del testo, alcuni *software* procedono inizialmente anche alla **lemmatizzazione** dei testi, che consiste nel riconoscimento automatico delle diverse flessioni di sostantivi, aggettivi, verbi, e nel riportarli alla loro radice comune.

La lemmatizzazione comporta che le forme dei **sostantivi** e degli **aggettivi** vengano ricondotte al **maschile singolare**, quelle dei **verbi** all'**infinito presente**, quelle delle preposizioni articolate alla loro forma senza articolo e così via.

2. Informazioni sui parlanti

- Oltre la descrizione dei contenuti del testo è possibile far interagire il testo stesso con le variabili disponibili sui parlanti
- Per valorizzare le potenzialità di analisi è quindi opportuno raccogliere i testi in modo che siano disponibili alcune variabili da associare ai parlanti (es. sesso, età, condizione professionale, data di produzione, istruzione etc.)

3. Primi passi nell'analisi: lo studio del vocabolario

L'esame delle parole diverse (o forme grafiche) che compongono il corpus è il primo passo da compiere nell'analisi testuale.

- Per mezzo del semplice conteggio delle parole viene creato il **vocabolario**, che consiste nell'elenco di tutte le differenti parole che compaiono nel corpus, utilizzato al fine di selezionare una serie di **parole significative** che consentano di interpretarne il contenuto.
- Le parole contenute nel vocabolario possono essere mostrate in ordine alfabetico oppure possono essere ordinate per valori decrescenti di frequenza.

Lo studio del vocabolario

- L’ordinamento alfabetico permette di individuare rapidamente eventuali errori di trascrizione dei vocaboli e di riconoscere parole contraddistinte dalla stessa radice (ad esempio ritornare, ritornata, ritornavo, ritornerò, tornare, torno, torniamo, etc.)
- L’ordinamento in senso decrescente di frequenza consente di evidenziare le parole che ricorrono più spesso nei testi analizzati.

Lo studio del vocabolario

Dal punto di vista semantico le parole di un testo non possono essere considerate equivalenti, si distinguono, infatti,

- le **parole vuote** che hanno un significato esclusivamente grammaticale stabilito solamente in relazione con altre parole. Si tratta di parole dal contenuto strumentale (di, e, che, per, ecc.), presenti in tutti i testi e poco informative (della Ratta-Rinaldi, 2007).

Più in generale, vengono definite come parole vuote, di volta in volta, tutte quelle parole che non riferiscono un contenuto rilevante ai fini dell'analisi (Giuliano, 2004).

Esempio di vocabolario con parole vuote

Tabella 2: Distribuzione di frequenza delle parole originarie nel corpus delle interviste alle donne rumene

Parole	Freq.
che	380
e	340
di	330
a	280
non	268
un	262
sono	250
ho	239
in	233
per	208
mi	185
la	183
perché	168
è	159
una	153
con	150
no	144
il	140
lavoro	**133**
si	126
i	119
ma	110
anche	105
Si	103
da	102
come	100
qui	96
io	**94**
anni	**93**
le	88
casa	**85**
l	84
ha	81
mia	80
più	79
se	68
ci	67
bene	67
po	65
adesso	**60**
Romania	**57**
mio	57
poi	55
era	55
quando	54
c	54
me	53
so	51
lo	49
al	48
molto	48
due	48
dopo	47
hanno	47
tempo	**44**
Italia	**44**
questa	44
lavorato	**44**
questo	44
nel	44
dei	43
persone	**43**
quindi	43
così	43
anno	**43**
dove	40
stata	40

Lo studio del vocabolario: le parole vuote

Per qualsiasi analisi del testo, indipendentemente dall'argomento trattato, l'esclusione delle parole vuote costituisce un passaggio obbligato, proprio in ragione della loro inconsistenza semantica.

Lo studio del vocabolario: le parole piene

- Una volta isolate le parole vuote, nel vocabolario si incontrano le cosiddette **parole piene** (o parole tema) ossia quelle parole ricche di significato, che contribuiscono significativamente all'interpretazione del testo, ne costituiscono l'ossatura fondamentale e che, proprio grazie alla loro elevata frequenza, consentono di distinguere immediatamente gli **argomenti**, i **contenuti** e i **protagonisti** principali del testo stesso.

- Dall'esplorazione dei risultati di questa preliminare analisi del vocabolario è possibile individuare le forme più significative in termini di occorrenze totali e, per estensione, i contenuti più frequenti, dunque considerati rilevanti

Esempio di vocabolario senza parole vuote

Tabella 6.2. Distribuzione di frequenza delle parole piene nel corpus delle interviste alle donne rumene

Parole	Frequenze	Parole	Frequenze
Lavoro	133	Tempo	44
Io	94	Italia	44
Anni	93	Lavorato	44
Casa	85	Persone	43
Adesso	60	Anno	43
Romania	57		

Le nuvole di parole

- La nuvola di parole è stata ottenuta ricorrendo a strumenti on line disponibili all'indirizzo http://www.tagcrowd.com

Rappresentazione del contenuto

- **Parole vuote**
- **Parole piene (o parole tema)** (alta frequenza nel testo)
- **Parole - chiave** (forme sovra o sotto rappresentate rispetto a un modello di riferimento)
- **Parole specifiche:** forme sovra o sotto rappresentate in alcune parti del testo

Lo studio del vocabolario

In realtà, questo tipo di approccio non è privo di difficoltà.

La più rilevante è determinata dal trattamento delle **ambiguità semantiche** delle unità di testo semplici.

In genere, il lavoro della **disambiguazione** è certamente il più gravoso e si effettua attraverso un continuo ritorno alla lettura dei testi nei contesti.

Frequenza delle parole

- Qualsiasi analisi del testo basata su aspetti statistici attribuisce un ruolo fondamentale alla frequenza delle parole, tuttavia questa condizione non va interpretata come **unica e sola condizione**, sia necessaria che sufficiente, per conferire valore alle parti di un vocabolario.

- Alcune parole usate con frequenza relativamente bassa, possono essere comunque rilevanti in virtù del contenuto che richiamano. Spesso una parola che compare una sola volta in un testo, può avere un ruolo fondamentale al servizio della comprensibilità del testo stesso.

- Dunque, è importante sottolineare la presenza di queste parole ed è ancora più importante includerle nell'analisi ed eventualmente classificarle in categorie.

4. Selezione dei segmenti ripetuti significativi

- Con l'obiettivo di approfondire i contenuti del testo è possibile ottenere selezioni ragionate del vocabolario attraverso lo studio dei **segmenti ripetuti.**

Si tratta di sequenze di parole nel testo, costituite da due o più forme, che riferiscono un **significato più preciso** rispetto alle singole parole, in quanto consentono di rendere più esplicite forme grafiche che prese singolarmente avrebbero minore significatività, di tenere conto dei contesti d'uso delle parole e di individuare, così, **attori**, **azioni** e **oggetti** attorno ai quali si articolano i testi in esame (della Ratta-Rinaldi, 2007).

4. Selezione dei segmenti ripetuti significativi

- attribuire importanza nulla ai segmenti composti da parole vuote (ad es. e con, e di, per la)
- estrarre automaticamente i segmenti più rilevanti (ad es. capo dello stato, tempo libero etc.)

5. Analisi dei contesti (o analisi delle concordanze)

- Fase di ritorno al testo originario per controllare il significato attribuito alle parole selezionate per l'analisi.
- Analisi dei contesti d'uso di una parola **X** effettuata visualizzando le n parole precedenti e le n parole successive, tutte le volte che la parola **X** compare nel testo.

Analisi dei contesti - I discorsi sulle DIFFICOLTÀ nella scrittura della tesi di laurea

*** Requête numéro 1 *** forme=difficoltà -> Fréquence = 839

ene seguito da qualcuno e ha grandissime	**difficoltà**	a capire come si scrive dove si va dove
livello burocratico cioè ho trovato più	**difficoltà**	a capire i tempi di consegna in superar
le di input di raccolta del materiale la	**difficoltà**	a capire il giudizio generale del profe
l impostazione che m ero cercato di dare	**difficoltà**	a conciliare i vari metodi a cui atting
sulla Telecom e ho trovato proprio tanta	**difficoltà**	a contattare gente tipo i professionist
senza andare al centro del problema, poi	**difficoltà**	a contattare i professori qui a Roma, a
nza alla dispersione nel senso una certa	**difficoltà**	a darmi proprio delle scadenze, fare pr
arretrati, oppure da studenti che hanno	**difficoltà**	a iscriversi ad altri corsi di studi, v
non hai nessuno a cui far presente delle	**difficoltà**	a lezione lotti per avere un posto a se
ne facciamo pochissimi forse	**difficoltà**	a livello di composizione non eccessive
i l ha corretta alla fine, quindi magari	**difficoltà**	a livello di incertezza se fosse impost
e qualcuno sa più di te dove andare, poi	**difficoltà**	a livello di informazioni pratiche dell
e in effetti è rimasta entusiasta quindi	**difficoltà**	a livello di tesi nessuna se devo esse
qualcosa e sono disponibili mah	**difficoltà**	a livello generale, perché essendo una
MISSING parecchie ho trovato	**difficoltà**	a livello personale di tempo perché sai
rocrazia interna insomma che ci mette in	**difficoltà**	a noi studenti perché non riusciamo poi
a ricerca empirica porta via molto tempo	**difficoltà**	a non finire ecco e quindi anche perché
lecentro di Roma Nexus ed ho avuto gravi	**difficoltà**	a ottenere i contatti cioè poter contat
che vuole fare. bho, la difficoltà la	**difficoltà**	a ottenere informazione dai poiché ho f
rda invece la ricerca sul campo ho avuto	**difficoltà**	a parlare soprattutto con una parte del
amo disponibile quindi non è che c avevo	**difficoltà**	a prenderci degli appuntamenti o a parl
sione degli argomenti, gli obiettivi, la	**difficoltà**	a reperire e a farti seguire dal profes
mancata l aspetto pratico però insomma.	**difficoltà**	a reperire il materiale essenzialmente,
teoria e poca pratica in sostanza	**difficoltà**	a reperire il materiale ma perché c era
organizzare il pensiero insomma. la	**difficoltà**	a reperire il materiale poi siccome la
fatto che non vivevo in Italia e quindi	**difficoltà**	a reperire il materiale principalmente
cui ho fatto la tesi e niente quindi la	**difficoltà**	a rintracciare queste persone gli sono
ivere in una bella forma perché io trovo	**difficoltà**	a scrivere bene cioè se prendo degli sp
sono state quelle più complesse proprio	**difficoltà**	a scrivere in italiano perché erano ann
i poi basta per il resto e diciamo anche	**difficoltà**	a scrivere in una bella forma perché io
nto che m interessava e poi inizialmente	**difficoltà**	a scrivere proprio. non ce ne sono stat
anche di fare dei corsi interessanti.	**difficoltà**	a strutturare l l impostazione della te
e è stato un incubo e poi ho avuto anche	**difficoltà**	a tradurre in linguaggio accademico que
tto alla data di pubblicazione già si ha	**difficoltà**	a trovare dei libri. però li abbiamo tr
è zero. dunque più che altro ecco	**difficoltà**	a trovare i riferimenti storici perché
rimenti storici perché ho avuto un po di	**difficoltà**	a trovare i testi e quindi più che altr

Un esempio: I discorsi sulle DIFFICOLTÀ nella scrittura della tesi di laurea

- le difficoltà nel reperimento del materiale di ricerca e la selezione della bibliografia;
- la scarsa assistenza ricevuta dal relatore;
- la necessità di circoscrivere gli argomenti e impostare il ragionamento;
- le difficoltà organizzative determinate dalla conduzione di una ricerca empirica.

6. Classificazione delle parole in categorie

Le parole piene e i segmenti ripetuti significativi che compaiono nel vocabolario possono essere successivamente classificati in **categorie** di diverso tipo:

- **Categorie semantiche**, nelle quali sono raggruppate parole ed espressioni che assumono, nell'unità di contesto, lo stesso significato o significati simili,
- e/o in **Categorie tematiche**, in ciascuna delle quali sono classificate parole ed espressioni che si riferiscono allo stesso tema o argomento.

Quindi, le categorie costituiscono i vocaboli del **dizionario** e definiscono campi semantici oppure tematici al loro interno omogenei in relazione a un particolare significato oppure a una determinata proprietà.

6. Classificazione delle parole in categorie

Le categorie sono concetti-chiave, etichette concettuali che raggruppano parole, locuzioni, frasi con significati analoghi oppure riferite a uno stesso argomento in un sistema di classificazione per argomento (Losito, 2002; 2007).

Questo tipo di analisi permette di rendere conto della **dimensione tematica** di un insieme di testi e di delineare una sorta di panoramica di sintesi dei **nuclei concettuali**, per mezzo dell'individuazione dei principali **elementi di contenuto** presenti nel corpus (della Ratta-Rinaldi, 2007).

Un esempio di applicazione: le interviste ai Giubilanti del 2000

In occasione del Giubileo del 2000, il gruppo di ricerca diretto da Roberto Cipriani ha svolto una serie di interviste a pellegrini provenienti da diverse aree geografiche (Cipriani 2002).

Sulla trascrizione di ciascuna di queste interviste il gruppo stesso ha effettuato un lavoro di codifica e di segmentazione del testo in funzione della presenza di riferimenti, nel testo stesso, a un insieme di items considerati significativi in funzione degli obiettivi conoscitivi della ricerca.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Un esempio di come sono stati codificati i testi delle interviste è il seguente:

*(*giubileo-inizio) (*emozione-inizio)* ***La mia ... la mia esperienza del giubileo è stata molto positiva perché è la cosa più bella che io ho avuto nella mia vita*** *(*giubileo-fine),* ***venendo da Taranto, perché abito a Taranto ... , la ...*** *(*comunità-inizio)* ***abbiamo fatto una bellissima gita sul pullman ... ridendo, scherzando ... con ... tutta la comitiva della confraternita,*** *(*Madonna-inizio)* ***perché noi a Taranto teniamo una confraternita ... chiamata la Madonna del Carmine*** *(*Madonna-fine)* ***... dopodiché*** *(*chiesa-inizio)* ***siamo arrivati a San Pietro ...*** *(*Roma-inizio)* ***Io già Roma lo conoscevo ... però ... per me ... è stata una gioia quando ho sentito che dovevamo venire a Roma*** *(*emozione-fine) (*Roma-fine) ...* ***entrando a ... a ... a San Pietro*** *(*chiesa-fine)…*

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Il primo passo compiuto è stata la traduzione di ciascuna intervista in un metatesto, che è stato costruito sostituendo a ciascun segmento del testo originario l'item ad esso corrispondente.

Il metatesto, dunque, consiste nella successione degli items nei quali ciascuna intervista è stata segmentata. Il brano prima riportato diventa pertanto:

Giubileo emozione comunità Madonna chiesa Roma

Un esempio di applicazione: le interviste ai Giubilanti del 2000

È stato necessario ricondurre i 91 items originari ad un numero ridotto di categorie tematiche, a ciascuna delle quali è stata associata un'etichetta verbale idonea a rappresentarne il contenuto.

Di seguito si riportano alcune delle categorie utilizzate (il *thesaurus*) con l'indicazione, per ciascuna categoria, di ciò che essa include:

Un esempio di applicazione: le interviste ai Giubilanti del 2000

- *Giubileo*: riferimenti al viaggio compiuto per giungere a Roma, all'esperienza del pellegrinaggio, alle motivazioni del viaggio, ai particolari relativi all'organizzazione e alla preparazione del viaggio e del soggiorno, ai disagi incontrati, alle aspettative nei confronti del Giubileo, alla Porta Santa, ad altri pellegrinaggi, al rito dell'indulgenza;

- *Emozione*: affermazioni riguardanti stati d'animo, sentimenti e, appunto, emozioni;

- *Roma*: riferimenti alla città ospite dell'evento;

- *Chiesa*: riferimenti alla Chiesa intesa come istituzione e ai diversi movimenti religiosi, eventuali critiche mosse alla Chiesa;

Un esempio di applicazione: le interviste dei Giubilanti del 2000

Sulla base di questo insieme di categorie, ciascun metatesto costruito nella fase precedente è stato trasformato in un secondo metatesto, costituito questa volta dalla successione delle etichette verbali associate alle categorie alle quali sono stati ricondotti gli items utilizzati nella prima trasformazione dei testi originari delle interviste.

Operando sul corpus costituito dall'insieme dei metatesti così predisposti il software ha fornito anzitutto due elenchi delle etichette verbali associate alle categorie, uno ordinato alfabeticamente e l'altro in base alle **frequenze** con cui le categorie stesse ricorrono.

Categorie	Items	Freq.
GIUBILEO	Viaggio, Pellegrinaggio, Motivazioni, Organizzazione, Disagi, Aspettative, Porta Santa, A Piedi, Indulgenza, Preparazione, Terra Santa	713
FEDE	Credenza, Dio, Gesù Cristo, Grazia, Stato di Grazia, Madonna, Vangelo, Sante (Sacre) Scritture, Anima, Miracoli, Spirito Santo, Speranza, Conversione, Desiderio, Conforto, Dono, Testimonianza, Ringraziamento	548
FAMIGLIA	Famiglia d'origine, Famiglia di Procreazione, Madre, Padre, Figli	492
VALORI	Amicizia, Amore, Etica, Tradizione, Patria, Volontariato, Solidarietà, Obbedienza, Impegno	357
RELIGIOSITA'	Pratica, Spiritualità, Pietà, Devozione, Culto, Rito, Perdono, San Francesco, Padre Pio, Preghiera	347
MALE	Morte, Malattia, Peccato, Stato di Peccato, Povertà, Solitudine	261
TEMPO	Età, Anziano, Giovani	193
SCUOLA	Educazione, Istruzione	185
ALTRI	Civiltà, Società, Comunità, Appartenenza	178
LAVORO		167
CHIESA	Movimenti Religiosi, Critiche alla Chiesa	163
ROMA		159
VITA		144
EMOZIONE	Animo	109
PAPA		105
MASS MEDIA	Messaggio, Internet, Modernizzazione	92
CHIESE	Luogo Fisico	90
POLITICA	Potere	50
SOLDI		45
REGOLE	Doveri	38
OPINIONE		36
ALTRE RELIGIONI		31
Totale nel corpus		**4503**

7. Le parole/categorie specifiche (o caratteristiche)

- Nell'ambito del modulo testuale delle versioni più recenti del *software* Spad la procedura ***VoSpec*** (*Vocabulaire Spécifique des Groupes d'individus*) consente di individuare, all'interno di un corpus di testi, le parole o le categorie di parole specifiche (o caratteristiche) di determinati gruppi di individui, avendo a disposizione informazioni sui produttori dei testi stessi.

7. Le parole/categorie specifiche (o caratteristiche)

L'analisi delle **parole caratteristiche** consente di analizzare un testo confrontando tra loro alcune partizioni.

Le **parole caratteristiche** sono quelle sovra-rappresentate o sotto-rappresentate in uno dei gruppi, e analizzarle consente di evidenziare il profilo lessicale (o tematico) caratteristico di ognuno di essi.

7. Le parole/categorie specifiche (o caratteristiche)

L'applicazione di tale procedura offre un ulteriore aiuto alla lettura critica dei testi, poiché il calcolo delle parole caratteristiche consente di esaminare il profilo di ciascuna categoria di produttori di testi, a partire da **un'ipotesi probabilistica di equidistribuzione** delle forme nel testo: quando una parola è **equidistribuita** in tutti i sotto-testi è considerata ***banale***; quando, invece, una parola risulta **sovrarappresentata** in uno dei sotto-testi è considerata ***caratteristica*** (o ***specifica***) (della Ratta-Rinaldi, 2007).

7. Le parole/categorie specifiche (o caratteristiche)

Con questa tecnica è possibile determinare il linguaggio o, più specificamente nel presente caso di studio, i **riferimenti tematici** sovrarappresentati in determinati gruppi di individui. Il vantaggio fondamentale dell'applicazione di questo tipo di analisi è dato dalla possibilità di far interagire le variabili extra-testuali con quelle testuali, per andare oltre la semplice **lettura sequenziale** dei testi e distinguere **differenze** altrimenti **difficilmente percepibili**.

Il risultato consiste, dunque, in un dizionario specifico (trattandosi di parole classificate in categorie) dei diversi gruppi stabiliti sulla base delle informazioni extra-testuali.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Categorie specifiche degli uomini intervistati

Categorie o segmenti caratteristici	% interna	% globale	Freq. interna	Freq. globale	Valore test	Probabilità
giubileo vita	0,86	0,42	13	19	2,871	0,002
giubileo giubileo giubileo giubileo	0,73	0,36	11	16	2,625	0,004
vita	4,17	3,20	63	144	2,509	0,006
religiosità lavoro	0,40	0,16	6	7	2,450	0,007
giubileo giubileo giubileo	1,66	1,09	25	49	2,368	0,009
altri giubileo fede	0,26	0,09	4	4	2,235	0,013
fede male giubileo	0,26	0,09	4	4	2,235	0,013
giubileo	17,48	15,83	264	713	2,100	0,018
lavoro fede	0,60	0,31	9	14	2,089	0,018
vita male	0,60	0,31	9	14	2,089	0,018

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Di particolare evidenza, per quanto riguarda gli uomini intervistati, è la specificità della categoria ***giubileo***. L'esperienza del pellegrinaggio è vista nei suoi diversi aspetti, con riferimenti ai momenti **dell'organizzazione, della preparazione del viaggio** per giungere a Roma e del **soggiorno**, ai **disagi** incontrati, alle **motivazioni** del viaggio stesso, alle **aspettative** nei confronti del Giubileo, alla Porta Santa, al rito dell'indulgenza. Testimonianza di ciò sono quei segmenti in cui la categoria *giubileo* è ripetuta più volte. Da sottolineare anche la categoria ***vita***, sia singolarmente che in associazione con *giubileo*, così come le sequenze ***religiosità lavoro*** **e** ***lavoro fede***. Esperienza giubilare, dunque, ma anche esperienza esistenziale e lavorativa. Sono inoltre da segnalare, per questo gruppo di intervistati, i riferimenti alla società, alla **comunità di appartenenza**, nonché quelli alla ***fede***, al credere, al trovarsi in uno stato di grazia, al richiedere una grazia o un dono, alla speranza, all'esperienza di conversione, al bisogno di conforto, alla testimonianza della fede. Da notare, infine, in relazione a *giubileo*, il segmento *fede male* e il segmento *vita male*, dove per ***male*** si intendono gli accenni alla morte, alla malattia, alla povertà, alla solitudine, al peccato e, più in generale, alla sofferenza.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Categorie specifiche delle donne intervistate

Categorie o segmenti caratteristici	% interna	% globale	Freq. interna	Freq. globale	Valore test	Probabilità
famiglia	11,83	10,93	354	492	2,710	0,003

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Solo la categoria ***famiglia*** risulta essere caratteristica delle donne intervistate. Nel raccontare di sé, le donne fanno riferimento in maniera preponderante sia alla famiglia d'origine, sia a quella costituita successivamente con il coniuge e i figli. Questo tema non è da considerarsi esclusivo, bensì **imprescindibile** rispetto a tutti gli altri. Infatti, prendendo in considerazione le categorie tematiche con valore test più basso di quello di *famiglia* (pari a 2,71) proprie di questo gruppo, si individuano argomenti quali *fede*, *giubileo*, *religiosità*, *male*, come singole categorie e in segmenti con altre.

8. Analisi delle corrispondenze lessicali

L'Analisi delle Corrispondenze Lessicali (ACL) applica ai dati testuali l'Analisi delle Corrispondenze (AC), una tecnica di analisi dei dati per variabili categoriali elaborata nell'ambito dell'approccio *Analyse des données* dalla scuola francese di J.P. Benzécri all'inizio degli anni Settanta. (Benzécri J.P., *L'analyse des données.* Dunod, Paris 1973.)

Trattandosi di un procedimento di tipo **fattoriale,** attraverso l'ACL è possibile individuare **dimensioni sottese ai dati** che sintetizzano le molteplici relazioni tra le variabili originarie costituite dalle parole (o dalle categorie di parole) presenti nel corpus in esame.

8. Analisi delle corrispondenze lessicali

Queste dimensioni sono dette **fattori**, mentre le variabili originarie sono dette **variabili attive**.

I fattori sono variabili di sintesi che riproducono la variabilità della matrice e possono rivelare dimensioni di senso latenti.

Spad permette, inoltre, di associare al testo **variabili extra-testuali (variabili illustrative)** come, ad esempio, quelle relative alle caratteristiche sociodemografiche, e di porre in relazione con queste le caratteristiche dei testi riscontrate mediante l'analisi (le forme lessicali o le categorie tematiche presenti nel testo).

Due modi di leggere i risultati dell'ACL

- **Interpretazione semantica**, basata sulla considerazione dei contributi, delle coordinate fattoriali e del valore test;

- **Interpretazione grafica**, basata sulla lettura del grafico fattoriale (proiezione sugli assi delle parole e/o categorie e delle variabili/modalità per mezzo delle quali sono stati ripartiti i testi)

L’interpretazione semantica dei risultati dell’ACL

L’interpretazione semantica dei fattori estratti si basa sui seguenti parametri:

- il **contributo assoluto** di ciascuna variabile attiva, che indica la quota di inerzia totale del fattore spiegata dalla variabile stessa, in altre parole, questo parametro rappresenta quanta parte ha avuto tale variabile nella determinazione del fattore, in rapporto all’insieme delle variabili;

- il **coseno quadrato**, che indica il contributo del fattore alla spiegazione della variabilità di una determinata variabile;

- Il **valore test**, che è un test di significatività statistica che indica se la relazione delle modalità delle variabili con i fattori è statisticamente significativa (per Ph_0=0,05, quando il V.T. ≥ 2).

L’interpretazione semantica dei risultati dell’ACL

Per individuare le modalità che maggiormente contribuiscono a generare assi e piani fattoriali, e che quindi danno loro significato, si selezionano per ciascun fattore le variabili e le relative modalità con **contributi assoluti più elevati.** Tra queste, si concentra l’attenzione sulle modalità che hanno i **coseni quadrati più alti.**

Un esempio di applicazione: le interviste ai Giubilanti del 2000

- Nell'analizzare le interviste dei pellegrini del Giubileo sono state utilizzate come variabili attive le etichette verbali associate alle **categorie** in cui sono stati classificati gli items originari;
- le variabili utilizzate, invece, come illustrative sono state **l'età**, codificata in tre fasce 18-35 anni, 36-60 anni, 61-75 anni, il **sesso** e il **gruppo linguistico** di appartenenza, codificato in "gruppo linguistico italiano" e "altri gruppi linguistici".
- Sono stati studiati i primi cinque assi fattoriali, che spiegano complessivamente il 49.62% dell'inerzia totale.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Istogramma degli autovalori

NUMERO FATTORE	VALORE PROPRIO	PERCENTUALE	PERCENTUALE CUMULATA	
1	**.1225**	**15.53**	**15.53**	**********
2	**.0852**	**10.80**	**26.33**	**********
3	**.0706**	**8.95**	**35.28**	**********
4	**.0584**	**7.40**	**42.68**	**********
5	**.0548**	**6.94**	**49.62**	**********
6	.0507	6.43	56.05	**********
7	.0431	5.47	61.51	**********
8	.0406	5.15	66.66	**********
9	.0339	4.30	70.96	**********
10	.0331	4.19	75.15	**********
11	.0296	3.76	78.91	**********
12	.0267	3.39	82.30	**********
13	.0248	3.14	85.44	**********
14	.0224	2.83	88.28	**********
15	.0193	2.45	90.73	**********

Un esempio di applicazione: le interviste ai Giubilanti del 2000

- Il primo fattore spiega il 15,53% dell'inerzia totale e mostra sul semiasse positivo l'associazione forte di items quali ***fede, religiosità*** **e** ***giubileo***, contrapposta sul semiasse negativo a categorie di carattere più spiccatamente sociale, quali ***famiglia, scuola, lavoro, soldi***.

- Le variabili illustrative mostrano come la modalità di raccontare l'esperienza giubilare, evidenziata sul semiasse positivo, sia riconducibile principalmente a intervistati del gruppo linguistico italiano, fino ai sessant'anni di età. Alle modalità rappresentate sul semiasse negativo si associano, invece, intervistati di gruppi linguistici diversi dall'italiano e di età più avanzata.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Primo fattore: Spirituale VS Sociale

SEMIASSE POSITIVO			SEMIASSE NEGATIVO		
Variabili attive	**Contributo assoluto**	**Coseno quadrato**	**Variabili attive**	**Contributo assoluto**	**Coseno quadrato**
Fede	8.5	0.20	Famiglia	28.8	0.61
Religiosità	8.1	0.28	Scuola	14.7	0.36
Giubileo	5.0	0.11	Lavoro	13.0	0.43
			Soldi	6.3	0.26
Variabili illustrative	**Valore test**		**Variabili illustrative**	**Valore test**	
Italiano	41.9		Altro gruppo linguistico	-41.9	
36-60 anni	6.3		61-75 anni	-9.2	
18-35 anni	3.1				

Un esempio di applicazione: le interviste ai Giubilanti del 2000

- Il secondo fattore, che spiega il 10,80% dell'inerzia totale, mostra sul semiasse positivo come nel racconto dell'esperienza giubilare si inseriscano riferimenti **all'esperienza** esistenziale di ciascuno oltre che alle **emozioni** vissute in relazione alla pratica religiosa, riferimenti espressi principalmente da donne, per lo più anziane, appartenenti al gruppo linguistico italiano.

- All'opposto, sul semiasse negativo il focus è centrato esclusivamente sull'evento **Giubileo**. Le variabili illustrative mostrano come questo modo di rappresentare l'esperienza del pellegrinaggio sia proprio dei giovani, dei maschi, e degli appartenenti ad altri gruppi linguistici.

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Secondo fattore: esperienza esistenziale VS esperienza giubilare

SEMIASSE POSITIVO			SEMIASSE NEGATIVO		
Variabili attive	**Contributo assoluto**	**Coseno quadrato**	**Variabili attive**	**Contributo assoluto**	**Coseno quadrato**
Tempo	13.3	0.31	Giubileo	49.0	0.74
Emozione	7.0	0.16			
Fede	6.7	0.11			
Male	7.3	0.17			
Vita	6.1	0.16			
Variabili illustrative	**Valore test**		**Variabili illustrative**	**Valore test**	
61-75 anni	20.9		18-35 anni	-23.0	
Italiano	17.6		Altro gruppo linguistico	-17.6	
Femmina	5.4		Maschio	-5.4	

L’interpretazione grafica dei risultati dell’ACL

L’ACL, inoltre, consente di rappresentare graficamente le associazioni tra parole e testi su piani delimitati da due assi fattoriali. Le coordinate fattoriali, con il segno + o -, indicano la posizione delle variabili sugli assi fattoriali e la loro distanza dall’origine degli assi stessi.

Gli assi fattoriali possono essere interpretati in qualità di dimensioni semantiche attraverso cui leggere il corpus: la vicinanza tra parole (o categorie di parole) sul piano fattoriale rinvia, infatti, a una loro combinazione o associazione nel testo, e l’esplorazione delle associazioni tra parole o categorie di parole contribuisce alla lettura/descrizione del corpus.

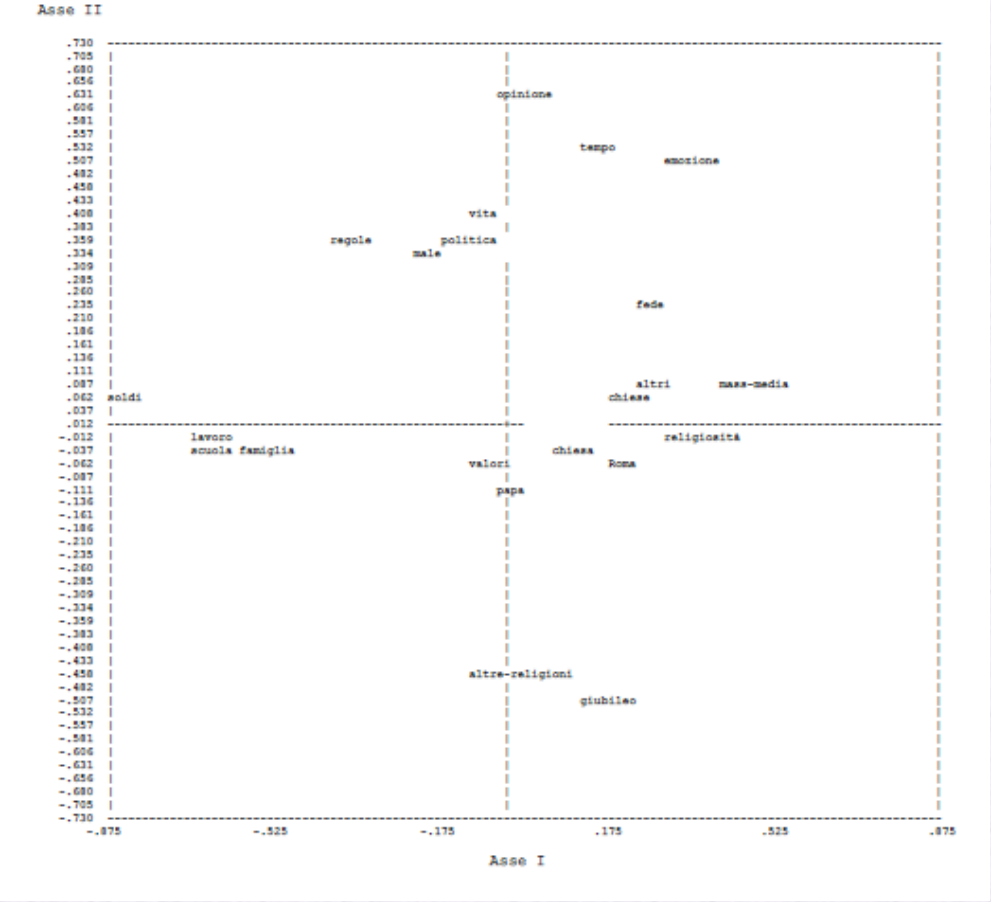
L’interpretazione grafica dei risultati dell’ACL

Per interpretare un fattore è utile analizzare le ‘opposizioni’ fra modalità rispetto ai semiassi positivo e negativo, cioè studiare attentamente i segni delle coordinate fattoriali delle modalità prese in considerazione, tenendo conto del fatto che, analizzando le intersezioni fra piani fattoriali, più un punto è lontano dall’origine di un asse, maggiore sarà il suo contributo alla formazione dell’asse stesso e più grande è la ‘vicinanza’ tra i punti, maggiore sarà la interdipendenza tra le modalità rappresentate da quei punti.

I segni positivo e negativo indicano semplicemente la posizione della variabile sugli assi fattoriali, non implicano in alcun modo una componente valutativa

Un esempio di applicazione: le interviste ai Giubilanti del 2000

Proiezione delle categorie sul piano fattoriale: assi I e II



Il risultato dell'ACL

Non si può affermare che l'analisi delle corrispondenze descriva in via definitiva i significati contenuti in un insieme di testi

Ma l'esplorazione delle associazioni tra le parole può far individuare alcune dimensioni di senso che possono contribuire alla lettura/descrizione del testo

In conclusione…

Sottoporre a questo tipo di analisi un corpus testuale, soprattutto un corpus che può essere molto vasto, consente di **andare oltre la lettura sequenziale dei testi** e distinguere **caratteristiche altrimenti difficilmente percepibili.**

Bibliografia essenziale

Aureli Cutillo E., Bolasco S. (a c. di), 2004, *Applicazioni di analisi statistica dei dati testuali*, Roma, Casa Editrice Università degli studi di Roma La Sapienza.

Bardin L., 1977, *L'analyse de contenu*, Paris, Presses Universitaires de France.

Bellelli G. (a c. di), 1989, *Il metodo del discorso. L'analisi delle produzioni discorsive in psicologia e in psicologia sociale*, Napoli, Liguori.

Benzécri J.P., 1973, *L'analyse des données,* Paris. Dunod.

Berelson B., 1952, *Content Analysis in Communication Research*, New York, The Free Press, (nuova ed. Hafner Publishing Co., 1971).

Berelson B., Lazarsfeld P.F., 1948, *The Analysis of Communication Content*, Chicago-New York, University of Chicago and Columbia University.

Bolasco S., 1999, *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*, Roma, Carocci.

Cannavò L., Frudà L. (a c. di), 2007, *Ricerca sociale*, 3 voll., Roma, Carocci.

Cecconi L. (a c. di), 2002, *La ricerca qualitativa in educazione*, Milano, Angeli.

Cipriani R. (a c. di), 2006, *L'approccio qualitativo. Dai dati alla teoria nell'analisi sociologica*, Milano, Guerini Scientifica.

Cipriani R., Bolasco S. (a c. di), 1995, *Ricerca qualitativa e computer. Teorie, metodi e applicazioni*, Milano, Franco Angeli.

Cipriani R., Losito G. (a c. di), 2008, *Dai dati alla teoria sociale. Analisi di un evento collettivo*, Roma, Anicia.

della Ratta-Rinaldi F., 2002, "L'analisi testuale, uno strumento per la ricerca qualitativa", in Cecconi L. (a c. di), *La ricerca qualitativa in educazione*, Milano, Angeli, 2002.

della Ratta-Rinaldi F., 2007, "L'analisi testuale computerizzata", in Cannavò L., Frudà L. (a c. di), *Ricerca sociale*, vol. II, Roma, Carocci, 2007, pp. 133-152.

della Ratta-Rinaldi F., 2007, "L'analisi multidimensionale dei testi", in Cannavò L., Frudà L. (a c. di), *Ricerca sociale*, vol. III, Roma, Carocci, 2007, pp. 133-150.

De Lillo A. (a c. di), 1971, *L'analisi del contenuto*, Bologna, Il Mulino.

De Sola Pool I. (ed.), 1959, *Trends in Content Analysis*, Urbana, University of Illinois Press.

D'Unrug M.-C., 1974, *Analyse de contenu et acte de parole*, Paris, Delarge.

Ercolani A.P., Areni A., Mannetti L., 2004[9], *La ricerca in psicologia*, Roma, Carocci (I ed. 1990).

Ericsson K.A., Simon H.A., 1984, *Protocol Analysis. Verbal Reports as Data*, Cambridge, MIT Press.

Fielding N.G., Lee R.M. (eds.), 1991, *Using Computers in Qualitative Research*, London, Sage.

Bibliografia essenziale

Gerbner G., Holsti O.R., Krippendorf K., Paisley W.J., Stone Ph. J., 1969, *The Analysis of Communication Content. Developments in Scientific Theory and Computer Techniques*, New York, Wiley.

Ghiglione R., 1995, "Teorie, metodi e uso del computer nell'analisi del contenuto: alcuni problemi", in Cipriani R., Bolasco S. (a c. di), *Ricerca qualitativa e computer. Teorie, metodi e applicazioni*, Milano, Franco Angeli, 1995, pp. 71-86.

Ghiglione R., Blanchet A., 1991, *Analyse de contenu et contenu d'analyse*, Paris, Dunod.

Giuliano L., 2004, *L'analisi automatica dei dati testuali. Software e istruzioni per l'uso*, http://www.ledonline.it, LED Edizioni Universitarie di Lettere Economia Diritto.

Giuliano L., La Rocca G., 2008, *L'analisi automatica e semi-automatica dei dati testuali*, Milano, LED Edizioni Universitarie di Lettere Economia Diritto.

Giuliano L., La Rocca G., 2010, *L'analisi automatica e semi-automatica dei dati testuali II strategie di ricerca e applicazioni*, Milano, LED Edizioni Universitarie di Lettere Economia Diritto.

Krippendorf K., 1980, *Content Analysis: An Introduction to its Methodology*, London, Sage (trad. it. *Analisi del contenuto. Introduzione metodologica*, Torino, Eri, 1983).

Lebart L., Morineau A., 1982, *SPAD Système Portable Pour l'Analyse des Données*. Paris, Cisia.

Lebart L., Morineau A., Bécue M., 1989, *Spad.T Système Portable Pour l'Analyse des Données Textuelles*, Paris, Cisia.

Lebart L., Salem A., 1988, *Analyse statistique des données textuelles. Question ouvert et lexicométrie*, Paris, Dunod.

Lebart L., Salem A., 1994, *Statistique textuelle*. Paris, Dunod.

Losito G., 2002[4], *L'analisi del contenuto nella ricerca sociale*, Milano, Franco Angeli (I ed. 1993).

Losito G., 2004, *L'intervista nella ricerca sociale*, Roma-Bari, Laterza.

Losito G., Piccini M.P., 2006, "L'analisi della dinamica discorsiva per la ricerca qualitativa", in Cipriani R. (a c. di), *L'approccio qualitativo. Dai dati alla teoria nell'analisi sociologica*, Milano, Guerini Scientifica, 2006, pp.62-83.

Losito G., 2007, "L'analisi del contenuto nella ricerca sociale", in Cannavò L., Frudà L. (a c. di), *Ricerca sociale*, Vol II, Roma: Carocci, 2007, pp. 117-132.

Bibliografia essenziale

Ogilvie D.M., Stone P.J., Kelly E.F., 1980, "Computer-Aided Content Analysis", in Smith R.B., Manning P.K. (eds.), *Handbook of Social Science Research Methods*, New York, Irvington, 1980.

Piccini M.P., 2008, "Discan, Spad.T e l'analisi qualitativa", in Cipriani R., Losito G. (a c. di), *Dai dati alla teoria sociale. Analisi di un evento collettivo*, Roma, Anicia, 2008, pp. 169-197.

Rosengreen K.E. (ed.), 1981, *Advances in Content Analysis*, London, Sage.

Rositi F., 1970, *L'analisi del contenuto come interpretazione*, Torino, Eri.

Rositi F., 1989, "L'amore folle fra l'analisi del contenuto e il computer", in Bellelli G. (a c. di), *Il metodo del discorso. L'analisi delle produzioni discorsive in psicologia e in psicologia sociale*, Napoli, Liguori, 1989, pp. 107-114.

Smith R.B., Manning P.K. (eds.), 1980, *Handbook of Social Science Research Methods*, New York, Irvington.